U.S. DEPARTMENT OF COMMERCE
National Technical Information Service
PB80-125099

Studies in the Review of the Photochemical Oxidant Standard

William F. Biller, East Brunswick, NJ

Prepared for

Environmental Protection Agency, Research Triangle Park, NC Office of Air Quality Planning and Standards

Jan 79

PB80-125099

United States
Environmental Protection
Agency

Office of Air Quality
Planning and Standards
Research Triangle Park NC 27711

EPA-450/5-79-004
January 1979

Air

EPA

Studies in the Review of the Photochemical Oxidant Standard

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before completing)

1. REPORT NO.

2.

3. RECIPIENT'S ACCESSION NO. PB 80 125099

4. TITLE AND SUBTITLE

STUDIES IN THE REVIEW OF THE PHOTOCHEMICAL
OXIDANT STANDARD

5. REPORT DATE January 1979, Date of Issue

6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)

William F. Biller

8. PERFORMING ORGANIZATION REPORT NO.

9. PERFORMING ORGANIZATION NAME AND ADDRESS

William F. Biller
68 Yorktown Road
East Brunswick, N.J. 08816

10. PROGRAM ELEMENT NO.

11. CONTRACT/GRANT NO.

68-02-2589

12. SPONSORING AGENCY NAME AND ADDRESS

Environmental Protection Agency
Research Triangle Park
North Carolina 27711

13. TYPE OF REPORT AND PERIOD COVERED

Final Report

14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

Studies were performed in three areas for the Office of Air Quality Planning and Standards to support its review of the photochemical oxidants standard: 1. the desirability of changing from a photochemical oxidants to an ozone standard, 2. the suitability of alternate ways of expressing the standard and, 3. the assessment of health risks at alternate levels of the standard. From the first study it was concluded that a change to an ozone standard could lead to benefits in administering the standard with little or no loss in protection of the public health and welfare. The second study concluded that national ambient air standards should be expressed in a statistical form. A statistical form is more in accord with the actual behavior of pollutant concentrations in the atmosphere. In the third area the mathematical methodology was developed for estimating the risk that health effect thresholds for sensitive populations will be exceeded during a given period of time at alternate levels of the ambient air quality standard. The method quantitatively treats the uncertainties in health effects thresholds and the extreme concentration levels of pollutant to be experienced in the given time period.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Air Pollution Oxidant Ozone Stantdard Risk		

18. DISTRIBUTION STATEMENT

Release unlimited

19. SECURITY CLASS (This Report) NA

20. SECURITY CLASS (This page) NA

21. NO. OF PAGES 173

22. PRICE A08-A01

EPA-450/5-79-004

Studies in the Review of the Photochemical Oxidant Standard

by

William F. Biller

68 Yorktown Road
East Brunswick, N.J. 08816

Contract No. 68-02-2589

EPA Project Officer: Michael Jones

Prepared for

U.S. ENVIRONMENTAL PROTECTION AGENCY
Office of Air, Noise, and Radiation
Office of Air Quality Planning and Standards
Research Triangle Park, North Carolina 27711

January 1979

ACKNOWLEDGMENTS

The author wishes to express his appreciation to Mr. Michael Jones of EPA, the Project Officer for the study, for his many useful suggestions and guidance. The author is particularly indebted to Mr. Thomas Feagans of EPA who is responsible for the underlying ideas on the risk assessment methodology which are given mathematical expression in this report. Acknowledgment and appreciation are also extended to Messers. Carl Nelson and Ted Johnson of PEDCo Environmental for their work on the application of the Weibull distribution to hourly average ozone concentrations which is basic to much of the work in this report.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS iii

LIST OF FIGURES vi

LIST OF TABLES viii

1.0 INTRODUCTION 1

2.0 CHANGE FROM A PHOTOCHEMICAL OXIDANTS TO AN OZONE STANDARD 2

- 2.1 Background 2
- 2.2 Alternative #1 - Remain with the Photochemical Oxidant Standards 3
- 2.3 Alternative #2 - Change to Ozone National Ambient Air Quality Standards 4
- 2.4 Conclusions and Recommendations 5
- 2.5 References 6

3.0 STATISTICAL FORMS OF THE PHOTOCHEMICAL OXIDANTS STANDARD 7

- 3.1 Background 7
- 3.2 Mathematical Treatment 10
- 3.2.1 Distribution of Hourly Average Concentrations of Ozone 10
- 3.2.2 Interrelationship of Concentration Level of Standard and Expected Exceedance Rate 16
- 3.2.3 Statistical Behavior of the Extreme Values of the Hourly Average Photochemical Oxidant Concentrations Observed Over a Given Period 19
- 3.3 Nonindependence of Ambient Concentrations of Pollutants 36
- 3.4 Comparison of Statistical and Deterministic Standards 47
- 3.4.2 Comparison on the Basis of Fractional Reduction in Precursor Emissions Required to Meet Standard 56
- 3.4.3 The Highest Second Highest Concentration 67
- 3.5 Statistical Standard Based on the Daily Maximum Hourly Average Concentration 72
- 3.5.1 Comparison of Hourly and Daily Maximum Hour Standards When Hourly Average Concentrations Exhibit No Dependence 73

TABLE OF CONTENTS (cont'd)

Page

3.5.2 Comparison of Hourly and Daily Maximum Standards When Hourly Average Concentrations Exhibit Dependence 80

3.5.3 Comparison of Deterministic Hourly and Daily Maximum Standards 88

3.6 Conclusions and Recommendations 91

3.7 References 95

4.0 ASSESSMENT OF RISK ASSOCIATED WITH POSSIBLE CHOICES OF A PRIMARY NATIONAL AMBIENT AIR STANDARD 96

4.1 Background 96

4.2 Underlying Principles of Risk Assessment Method 96

4.3 Mathematical Description of the Method 100

4.4 Obtaining the $P_C(C)$ Distributions 107

4.5 Determination of P_C Functions for Ozone 118

4.6 Example Application of Method 124

4.7 Conclusions and Recommendations 129

4.8 References 131

APPENDIXES

APPENDIX A: SUPPORT MATERIAL FOR SECTION 3 132

APPENDIX B: DERIVATION OF BASIC EQUATIONS FOR ASSESSING HEALTH RISKS ASSOCIATED WITH ALTERNATIVE AIR QUALITY STANDARDS 136

APPENDIX C: EXAMPLES OF COMPOSITE HEALTH EFFECT THRESHOLD DISTRIBUTIONS FORMED FROM SIMPLE HYPOTHETICAL DISTRIBUTIONS FOR INDIVIDUAL THRESHOLDS 146

APPENDIX D: DETAILS OF CALCULATIONS PERFORMED IN SECTION 4.5 EXAMPLE APPLICATION OF METHOD 152

LIST OF FIGURES

Figure Page

3-1 Three Weibull Distributions with Expected Exceedances per Year of One at a Concentration of 0.1 ppm. 15

3-2 Effect of Weibull Distribution Shape Factor on Distribution of Highest Value. 27

3-3 Probability Density Function of Annual Highest Concentration for Different Values of Weibull Distribution Shape Factor. 29

3-4 Distribution of Annual Extremal Values for a Weibull Distribution of Hourly Concentrations Assuming Independent Hours 30

3-5 Probability Density Functions for Annual Extremal Values for a Weibull Distribuion of Hourly Average Concentrations (Assumes Independence of Hours). 31

3-6 Probability of a Given Observed Number of Exceedances as a Function of the Expected Exceedance Rate. 50

3-7 Probability of a Actual Exceedances Equal to or Above a Given Number as a Function of the Expected Exceedance Rate. 51

3-8 Distribution of Second Highest Value for Exponential Distribution of Hourly Average Concentrations. 60

4-1 Hypothetical Threshold Probability Density Function for a Health Effect with a 20% Chance of Being Real. 104

4-2 Variation in Composite Health Effect Threshold Probability Density Functions of Two Independent Density Functions as Different Probabilities of Existence are Assigned to the Lower Threshold Function. 106

4-3 Hypothetical Distribution of Hourly Average Concentration and the Corresponding Distribution of the Annual Maximum Hourly Average Concentration. 112

4-4 Change P_C Function with Change in Standard Concentration Level. 113

4-5 P_C Function for One Year Period for Different Values of Weibull Scale Factor, k. 114

4-6 P_C Function for 5 Year Period for Different Values of Weibull Scale Factor, k. 115

LIST OF FIGURES (cont'd)

Figure		Page
C-1	Composite Threshold Probability Densities When Health Effects Are Certain to Exist.	148
C-2	Composite Threshold Probability Densities When the Existence of One or More Health Effects is Not Certain.	149
D-1	Change in Function $P_C(C)$ with Change in C_{STD} for Exponential Distribution.	153
D-2	Graphical Illustration of p_T Calculation.	159

LIST OF TABLES

Table		Page
3-1	Levels of Standards with Expected Exceedances per Year Equal to One Which are Equivalent to a Standard with a Level of 0.08 ppm but with Expected Exceedances per Year Greater than One.	17
3-2	Levels of Standards with Expected Exceedances per Year Greater than One Equivalent to Standard with Level of 0.08 ppm and Expected Exceedance per Year Equal to One.	18
3-3	% Risk of m Exceedances/yr for E Expected Exceedances	22
3-4	% Risk of m or More Exceedances/yr of a Given Standard Level for E Expected Exceedances.	23
3-5	Relative Positions of Concentrations Corresponding to Various Expected Exceedance Rates for Different Weibull Distributions.	32
3-6	Location of the Most Probable Value of mth Highest Concentration as a Function of Number of Hours, n and Weibull Shape Factor, k.	33
3-7	Location of Expected Value of the mth Highest Concentration for an Exponential Distribution Relative to the Concentration with an Expected Exceedance Rate m for Two Levels of n.	35
3-8	Comparison of Distributions of Annual Maximum Hourly Average Concentration as Determined by Two Related Hourly Distributions.	41
3-9	Effective Increase in Weibull Constant k Needed to Give Same Distribution of Maximum Value as n is Decreased - Base Case: k = 1.0.	43
3-10	Effective Increase in Weibull Constant k Needed to Give Same Distribution of Maximum Value as n is Decreased - Base Case: k = 0.5.	44
3-11	Change in Weibull Distribution Shape Factors with Changing Time Segment in Which Hourly Average Concentrations are Collected.	45
3-12	Probability of a Given Observed Number of Exceedances for Different Expected Exceedance Rates.	52
3-13	Comparison of Statistical Forms with a Deterministic Standard Which Allows No More than One Exceedance per Year.	54

LIST OF TABLES (cont'd)

Table		Page
3-14	Cumulative Distribution of Second Highest Concentrations for 60 AQCR's.	58
3-15	Location of Most Probable Value of Second Highest Concentration Relative to the Concentration Whose Exceedance Rate is Two (C_2).	61
3-16	Change in Percent Control Required to Meet Standard at 0.08 ppm Level with Change in Allowed Expected Exceedances.	62
3-17	Change in Percent Control Required to Meet Standard at 0.08 ppm Level with Change in Allowed Expected Exceedances.	63
3-18	Change in Percent Control Required to Meet Standard at 0.08 ppm Level with Change in Allowed Expected Exceedances.	64
3-19	Change in Percent Control Required to Meet Standard at 0.08 ppm Level with Change in Allowed Expected Exceedances.	65
3-20	Change in Percent Control Required to Meet Standard at 0.08 ppm Level with Change in Allowed Expected Exceedances.	66
3-21	Position of Most Probable Highest Second High Concentration of N Years Relative to Concentration with Two Expected Exceedances per year.	70
3-22	Density Function for Distribution of Highest Second Highest Value in N Trials of 8760 Trials Each for Weibull Distribution of Hourly Averages Where $k = 1$, $C_E = 1.0$, $E = 2$.	71
3-23	Comparison of Daily Maximum Hourly Average and Hourly Average Standards.	78
3-24	Comparison of Daily Maximum Average and Hourly Average Standards with Partial Correction for Nonstationarity in Hourly Average Concentrations.	79
3-25	Distribution of Annual Six Highest Hourly Average Concentration by Region.	81
3-26	Daily Maximum Hourly Average Standard versus Hourly Average Standard.	83
3-27	Distribution of Exceedances per Year for Daily Maximum Hour Standard versus Exceedances per Year for Hourly Standard	85

LIST OF TABLES (cont'd)

Table		Page
3-28	Approximate Comparison of Statistical Daily Maximum Hour and Hourly Standards When Both Have Expected Exceedances per Year of One.	87
3-29	Comparison of Percent Hydrocarbon Control Required by Daily Maximum Hourly Average and Hourly Average Standards.	89
4-1	Subjective Probability Distributions of Lower and Upper Bounds of Weibull Shape Factor k for n at Approximately 1000 Hours and Maximum Concentration at Approximately 0.1 ppm.	121
4-2	Subjective Probability Distributions of Lower and Upper Bounds of Weibull Shape Factor k For n at Approximately 1000 hours and Maximum Concentration at Approximately 0.1 ppm.	122
4-3	Subjective Probability Distributions for Upper and Lower Bound Weibull Shape Factors (k) for Distributions of Hourly Average Ozone Concentrations.	123
4-4	Risks of an Hourly Average Concentration Exceeding True Health Effect Thresholds in a Given Year for Different Values of a National Ambient Air Quality Standard.	127
D-1	Individual and Composite Health Effects Threshold Probability Density Functions Vs. Concentration.	155

STUDIES IN THE REVIEW OF THE PHOTOCHEMICAL OXIDANTS NATIONAL AMBIENT AIR QUALITY STANDARDS

1.0 INTRODUCTION

The Clean Air Act requires that the U.S. Environmental Protection Agency periodically review the five National Ambient Air Quality Standards (NAAQS). The Office of Air Quality Planning and Standards is responsible for conducting this review and initiated the first round with a reconsideration of the standard for photochemical oxidants. A complex schedule of activities had to be completed in reviewing the photochemical oxidants standard and required the analysis of a number of issues. The contractor contributed to the analyses performed in three areas: 1. the desirability of changing the photochemical oxidants standard to an ozone standard; 2. the suitability of alternate ways of expressing the oxidant standard; and 3. the assessment of risk in setting the standard. The last analysis combined the considerations underlying statistical forms of the standard developed in the second analysis with the consideration of the uncertainty in the location of health effect thresholds. The fundamental idea of combining these two considerations to estimate risk is due to Mr. Thomas Feagans of The Office of Air Quality Planning and Standards. The contractor's contribution was development of the mathematical formalism which followed from the concept and which could be used to make actual risk estimates.

The report is divided into three major sections covering the areas listed above. Each section ends with conclusions and recommendations for the study covered.

2.0 CHANGE FROM A PHOTOCHEMICAL OXIDANTS TO AN OZONE STANDARD

2.1 Background

When primary and secondary national ambient air quality standards were first established for photochemical oxidants, it was proposed that the potassium iodide (KI) method be designated as the Federal Reference Method (FRM) for measuring photochemical oxidants. The method had been widely used for many years in making aerometric measurements of photochemical oxidants and measuring photochemical reactions in laboratory irradiation chambers. However, concern over the reliability and accuracy of the KI method led to the adoption of the newly developed chemiluminescent method as the FRM[1].

The chemiluminescent method is specific for ozone which is the major component of photochemical oxidants. It does not measure the remaining oxidants, the most prominent of which is peroxy acetyl nitrate (PAN). The inability to measure all, or nearly all of the oxidants present is an undesirable feature in an FRM for photochemical oxidants. The decision to use a measurement method specific for ozone was made necessary by the absence of satisfactory alternative methods. Furthermore the prospects are poor for developing a satisfactory photochemical oxidants measurement method. There are a large number of individual compounds that potentially could be classified as photochemical oxidant. To date only a relatively small number have been identified in the atmosphere or in laboratory irradiation chambers. Developing a practical measurement technique that would elicit a uniform response from such a complex class of compounds would be very difficult.

This discrepancy between the standard and the FRM has raised the question as to whether the present standard is not, in effect, an ozone standard because of the FRM and has led to suggestions that the standard be changed to ozone.

Two alternatives were considered in the present review of the photochemical oxidants standards: (1) Remain with the present standards and FRM with ozone as a surrogate for photochemical oxidants, and (2) Change to ozone standards. These alternatives are discussed in the following paragraphs.

2.2 Alternative #1 - Remain with the Photochemical Oxidant Standards

Remaining with the present standards has the advantage of continuing to emphasize the importance of protecting against the undesirable effects of all photochemical oxidants, ozone as well as the non-ozone oxidants.

The major difficulty is that ozone has problems as a surrogate for photochemical oxidants. It is not an accurate indicator of the total oxidants present and, perhaps more important, it provides little indication of the non-ozone oxidants present in the ambient air. Ambient ozone concentrations can vary from somewhat less than two-thirds to greater than 99% of the total photochemical oxidants concentration. The ratio of PAN (the second most predominant photochemical oxidant) to ozone can vary from 1/3 to 1/150[2] from one geographic location to another. It is likely that the ratios of ozone to other photochemical oxidants also vary over wide ranges.

Another difficulty with the present standard is that, aside from ozone, not much is known about the health and welfare effects of photochemical oxidants. Examination of the newly revised criteria document[1] for photochemical oxidants shows the majority of studies to be concerned with the effects of ozone. While early epidemiological studies which were the basis for the present standard[3] used the KI method (which is not a reliable measure for total oxidants) to estimate total oxidants, more recent studies tended to emphasize the use of chemiluminescence and other methods which are specific for ozone.

2.3 Alternative #2 - Change to Ozone National Ambient Air Quality Standards

A change to ozone standards would resolve the major problems with the present standards. It also would have the added features that the Federal Reference Method and a widespread ozone monitoring network are already in place. Ozone is a major atmospheric pollutant in its own right. It is the dominant component in photochemical oxidants (60 to 99%); and there is a large and growing body of information on its health and welfare effects. As indicated in the previous section, the new criteria document[1] is primarily concerned with ozone; the clinical and toxicological studies discussed in the document deal almost exclusively with ozone.

While a change to ozone standards would result in a loss of overt emphasis on controlling photochemical oxidants as a class, it is unlikely to result in increased risk of adverse effects to the public from photochemical oxidants or other products from atmospheric photochemical reactions of ozone precursors. A study of the new criteria document indicates that the principle health effects associated with photochemical oxidants are also caused by ozone alone and may be due primarily to the presence of ozone. For example, the epidemiological study by Toyama and Kagawa in which both ozone and photochemical oxidants were measured showed that ozone concentrations correlated more strongly with observed health effects than did total oxidant as estimated by the KI method[1].

A national ambient air quality standard for ozone would be based upon the effects of ozone in the ambient air. This means that estimations of effects thresholds and margin of safety considerations would take into account, to the extent allowed by available information, additive and interactive effects with non-ozone oxidants and other compounds present.

Finally, precursor emission control measures for the reduction of ozone in the atmosphere would also generally reduce the levels of the non-ozone oxidants and other compounds generated by the photochemical reactions of the precursors[4]. For example, studies suggest that hydrocarbon control measures designed to reduce ozone would produce a larger reduction in PAN[5]. Thus, there will be a reduction not only in the contribution from non-ozone oxidants and related photochemical pollutants to effects primarily caused by ozone, but also to effects, such as eye irritation and visibility reduction, which are largely due to non-ozone oxidants and other photochemical products. At the anticipated concentration levels of an ozone standard, these latter effects should be adequately controlled, based on present information.

It is expected, therefore, that ozone ambient air standards would provide the same degree of protection as photochemical oxidants standards. Futhermore, it would not be necessary to make any changes in regulatory strategies or programs as a direct result of a change from the one standard to the other. This latter conclusion also follows in part from the fact that current strategies and programs are already based on reduction of ozone levels in the atmosphere.

2.4 Conclusions and Recommendations

Given: 1. the difficulties with the present photochemical oxidants standards; 2. the suitability of ozone as a basis for air quality standards, 3. the equivalent protection provided by an ozone standard, based on present knowledge, and 4. the absence of any impact on regulatory strategies and programs, it is concluded that changing from the present primary and secondary photochemical oxidants standards to ozone national ambient air quality standards could lead to benefits in administering the standard and little or no loss in protection of the public health and welfare.

2.5 References

1. "Air Quality Criteria for Photochemical Oxidants and Oxidant Precursors", Vols. I and II, U.S. Environmental Protection Agency, Research Triangle Park, N. C. 27711.

2. "Analysis of Photochemical Oxidant Mix in the United States", Report to EPA in preparation, PEDCo Environmental, Inc.

3. "Air Quality Criteria for Photochemical Oxidants", Report No. AP-63, U.S. Department of Health, Education, and Welfare.

4. "Control of Photochemical Oxidants - Technical Basis and Implications of Recent Findings", U.S. Environmental Protection Agency, Office of Air Qualtiy Planning & Standards, Research Triangle Park, N.C. 27711, July 15, 1975.

5. W.A. Lonneman, J.J. Bufalini, and R.L. Seila, "PAN and Oxidant Measurement in Ambient Atmospheres", Environmental Science and Technology, Vol. 10, No. 4, 374 (1976).

3.0 STATISTICAL FORMS OF THE PHOTOCHEMICAL OXIDANTS STANDARD

3.1 Background

The present short-term National Ambient Air Quality Standards (NAAQS) established by the Environmental Protection Agency all specify a concentration level which is not be exceeded more than once in any year at a given air monitoring site. Compliance with the standard is simply determined by counting the number of times the specified concentration level is exceeded in a calendar year at a given monitoring site. If the number is greater than one the Air Quality Control Region (AQCR) in which the monitor is contained is not in compliance. Alternatively, the second highest concentration (averaged over the number of hours specified by the standard) observed during the year at each monitoring site is recorded. If any of the second highest values are above the standard level, the level has been exceeded more than once at some site and the AQCR is not in compliance. The highest second high concentration of the pollutant for all the monitoring sites within an area is considered to be a measure of the air quality over the area relative to that pollutant. This number can also be used as a "design value" to calculate how much emission reduction is needed to meet the standard. For photochemical oxidants proportional rollback, Appendix J[1], or the recently developed Empirical Kinetic Modeling Approach[2] are used in performing this calculation.

A single exceedance of the specified concentration level was provided for in the standard to allow for the occasional occurrence of weather conditions and/or emission conditions leading to an unusually high peak pollutant level and also to allow for an occasional monitoring error. However, there are problems with attempting in this manner to take into account the random changes in atmospheric pollutant concentrations caused by random fluctuations in weather and in emissions. For example, if certain meteorological conditions leading to concentrations above the standard can occur once in a given year, there is a definite probability they can occur more than once in other years and, therefore, lead to violations of the standard. It follows that the present deterministic standards can only be complied with on a continuing basis if the specified concentration level is never exceeded. Thus, the standard could not be exceeded even in years in which very rare adverse weather conditions occurred. Enforcing compliance under such rarely occurring conditions can require emissions to be reduced to levels far below levels that could be tolerated in more typical and frequently occurring years.

Another major difficulty is that the deterministic form of the present standards does not lend itself to the development of meaningful measures of existing air quality over an AQCR which can then be compared with the standard. As indicated above, the present practice is to use the annual second highest concentration measurement. However, the second high value is itself a random variable and will vary in a given area from year to year even though the average level of emissions leading to the observed atmospheric concentrations of the pollutant remains constant. This situation can result in uneven treatment from one AQCR to the next.

Present EPA guidelines recommend that several years of data be obtained and that the highest observed second high value be used. This guidance is in keeping with a strict interpretation of the deterministic standards and gives a more stable number than the second highest value from a single base year. However, the highest second high of a small number of years is still a random variable and, therefore, subject to change with fluctuations in weather and emissions.

The way out of these difficulties is to express the standard in a statistical form. For example, a statistical form that would result in the least departure from the present expression for photochemical oxidants is:

> 160 $\mu g/m^3$ (0.08 ppm) hourly average concentration with an expected number of exceedances per year less than or equal to one.

By setting the maximum permitted expected or average number of exceedances per year at one, allowance is made for occasional years in which the standard level is exceeded more than once so long as the average number of exceedances per year in not greater than one.

Furthermore, the statistical standard provides a meaningful method of comparing existing air quality with the standard. That concentration which has an expected exceedance rate of once per year at the existing air quality level can be compared directly with the standard level. This concentration, which can be designated as C_1, would be determined for each monitoring site within the AQCR and the highest number used as a design value for calculating the degree of emissions reductions needed to attain the standard. This follows because the concentration, C_1, unlike the second high concentration, is not a random variable. Instead C_1 has a single value corresponding to the average level of the emissions giving rise to the observed atmospheric concentration of the pollutant.

The value of C_1 for an AQCR would be estimated from one or more years of data. The estimate, of course, could fluctuate as the data base is expanded or reduced. However, the value of C_1 can be estimated from the overall distribution of hourly average concentrations, rather than determined as a single value as with the second high value. The estimate of C_1 is less subject, therefore, to random fluctuations than the second high value. It is also less subject to experimental error and less influenced by missing values than the second high concentration.

Other statistical expressions could also be used. For example, the standard could specify the maximum permitted value of the expected maximum one hour average concentration in one year, or it might specify the maximum permitted probability that a certain one hour concentration level will be exceeded in a given year. However, estimating the appropriate statistical measure of air quality in these cases is more complex than for the standard based on an expected exceedance rate.

The contractor under EPA Contract No. 68 02 2393 investigated the relative merits of deterministic and statistical standards[3]. These studies were extended under the present contract. The results are discussed in the following paragraphs.

3.2 Mathematical Treatment

Before proceeding to discuss the various aspects of statistical standards studied, it will be worthwhile to put in one place the mathematics used in the treatment and to discuss the behavior of measured atmospheric concentrations of ozone (present monitoring instruments predominately measure ozone).

3.2.1 Distribution of Hourly Average Concentrations of Ozone

The distribution of hourly average ozone concentrations over one year may be defined as follows:

$$F(C) = \Pr(C_{obs} \leq C) \qquad (3\text{-}1)$$

where F(C) may be considered the distribution function, Pr refers to the probability of the relationship within the brackets; C_{obs} is an observed hourly average concentration, and C a continuous variable representing hourly average concentration. The probability density function associated with F(C) is by definition:

$$f(C) = \frac{dF}{dC} \tag{3-2}$$

It has the property that:

$$\int_{C_1}^{C_2} f(C)dC = Pr(C_1 \leq C_{obs} \leq C_2) \tag{3-3}$$

Because the standard is concerned with the behavior of the highest hourly average concentrations in any one year a more conveneient form of the distribution function is:

$$G(C) = Pr(C_{obs} > C) \tag{3-4}$$

It follows from the theory of probability that:

$$G(C) = 1 - F(C) \tag{3-5}$$

Strictly speaking a $\geq$ sign should be used in Eq. (3-4), but since the function is defined as continuous there is no loss in accuracy in using the above more convenient definition.

Studies by Larson[4] and others have shown that the distributions of short-term averages concentrations of atmospheric pollutants are described reasonably well by lognormal distributions. It had been observed, however, that the lognormal distribution often did not give good fits on the high concentration end to the photochemical oxidants hourly average concentrations. Curran and Frank[5] suggested that exponential and Weibull distributions were useful alternatives to the lognormal distribution. Studies by PEDCo Environmental, Inc.[6] of oxidant/ozone air monitoring data from a number of locations confirmed that the Weibull distribution gave a better fit than lognormal in most cases. The Weibull distribution expressed in the form of Eq. (3-4) is:

$$G(C) = e^{-(C/\delta)^k} \qquad (3\text{-}6)$$

The equation has two parameters, δ and k. The parameter δ is called the scale factor. It is the concentration corresponding to $G(C) = 0.368$. It, therefore, tends roughly to place the midportion of the distribution. The parameter k is called the scale factor. It tends to be a measure of the spread of the distribution. The larger k, the more compact the distribution. PEDCo found that k values vary from one region to the next and may appear to vary within an area from one year to the next. For a full year of data (8760 hours) the observed range is about $k = 0.5$ to 2.3. The exponential distribution referred to above is a special case of the Weibull distribution obtained by setting $k = 1$. In this case it can be shown that δ becomes the mean concentration of the distribution.

The density function is found by differentiating Eq. (3-6):

$$f(C) = (k/\delta)\,(C/\delta)^{k-1}\,e^{-(C/\delta)^k} \qquad (3\text{-}7)$$

Equations (3-6) and (3-7) can be put in a form that is particularly convenient for dealing with a statistical standard which specifies a concentration level with a specified expected exceedance rate. From Eq. (3-6) it can be written that:

$$\frac{E}{n} = e^{-(C_E/\delta)^k} \tag{3-8}$$

Where E is the average or expected number of times the concentration C_E will be exceeded in n hours. Because, in the case of photochemical oxidant concentrations each hour in the year is not equivalent to every other hour, n is restricted to be a multiple of the number of hours within the year to which the distribution applies. Normally, this would be a multiple of 8760, the total hours in a 365 day year.

E is the expected number of exceedances in n hours because the probability, G(C), can be interpreted as the expected frequency of occurrence of hourly average concentrations above C that will be approached as the number of hours increases without limit. If n is the number of hours in one year and actual exceedances are observed for a large number of years their average value would therefore approach E in Eq. (3-8) as the number of years increased without limit. The quantity E is, therefore, the expected exceedance rate corresponding to n hours in a year.

Continuing with the development of more convenient forms of Eq. (3-6) and (3-7): Eq. (3-8) can be solved for the scale factor δ:

$$\delta = C_E\left(\ln\frac{n}{E}\right)^{-\frac{1}{k}} \tag{3-9}$$

and the Weibull distribution, Eq. (3-6) can be written:

$$G(C) = e^{-(\ln\frac{n}{E})} (C/C_E)^k \tag{3-10}$$

The corresponding density function becomes:

$$f(C) = \frac{k}{C_E}(\ln\frac{n}{E}) (C/C_E)^{k-1} e^{-(\ln\frac{n}{E})} (C/C_E) \tag{3-11}$$

Figure 3-1 shows three Weibull distribution functions. For each curve $C_E = 0.1$ ppm, $E = 1$, $n = 8760$. Their shape factors vary from 0.5 to 2.0. Over this range the degree of tailing in the high concentration range is seen to vary substantially. It is also noted that plotted on a semilogarithmic paper the Weibull distribution with $k = 1$ (exponential distribution) plots as a straight line. Special graph papers can be constructed which will give a straight line for any value of k (log log 1/G vs. C[6].

A distribution of hourly averages with a shape factor k just meeting a standard expressed as

> C_{STD} μg/m^3 (or ppm) hourly average concentration with the number of expected exceedances per year less than or equal to E_{STD}.

would then have the form:

$$G(C) = e^{-(\ln\frac{n}{E_{STD}})} (C/C_{STD}) \tag{3-12}$$

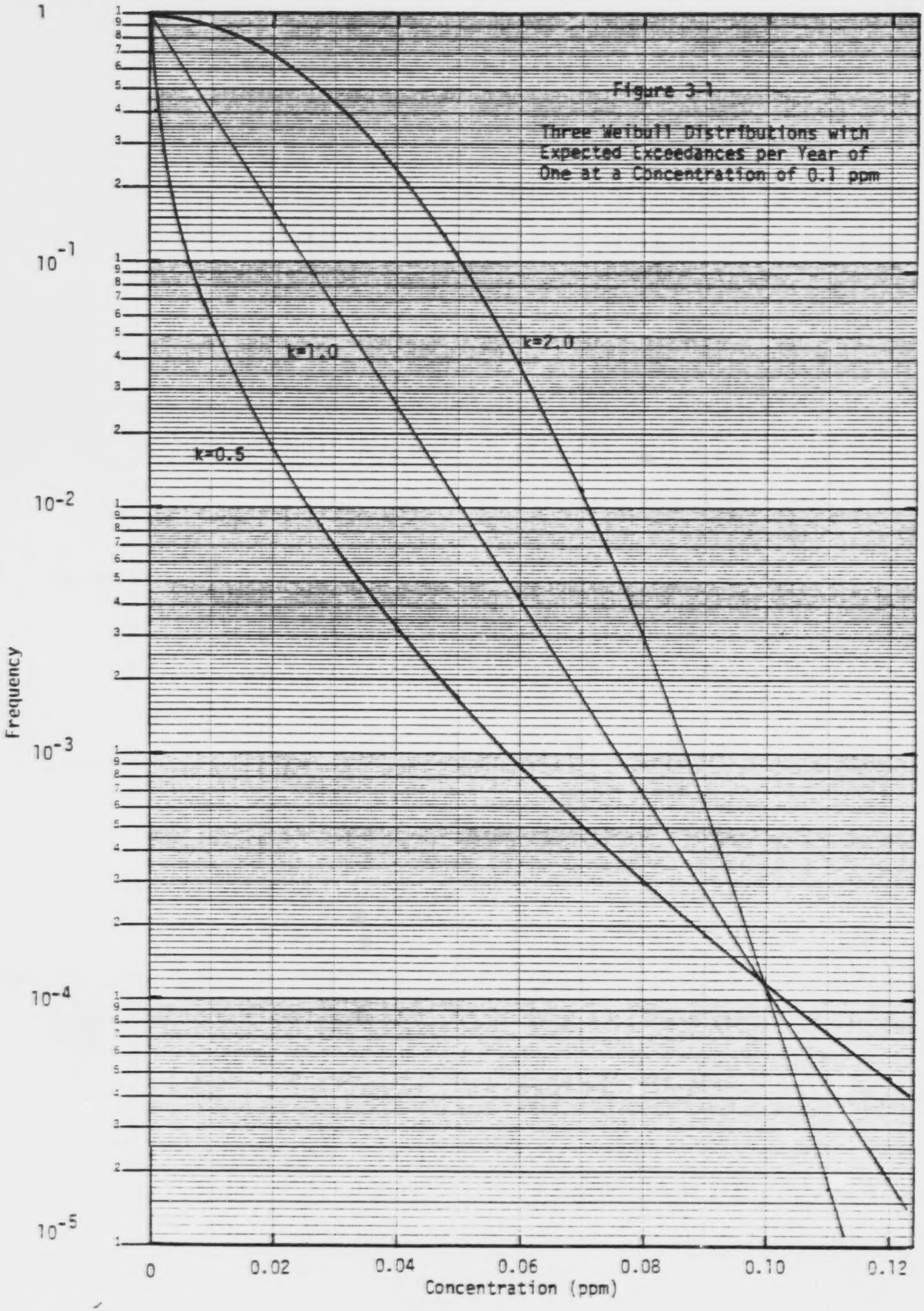


Figure 3-1

Three Weibull Distributions with Expected Exceedances per Year of One at a Concentration of 0.1 ppm

3.2.2 Interrelationship of Concentration Level of Standard and Expected Exceedance Rate

Given the statistical form of the standard shown in the last section it would appear that there is a trade-off relationship between the specified concentration level, C_{STD}, and the expected exceedance rate E_{STD}. That is, E_{STD} could be increased and C_{STD} lowered while maintaining the same level of protection . For example, if it is assumed that the Weibull function Eq. (3-10) adequately describes the distribution of hourly average concentrations over a monitoring site and the standard concentration level and expected exceedance rate are given as (C_{STD}, E_{STD}). The area around the monitoring site is just in compliance if:

$$\frac{E_{STD}}{n} = e^{-(\ln\frac{n}{E})\,(C_{STD}/C_E)} \qquad (3\text{-}13)$$

or

$$\frac{C_E}{(\ln\frac{n}{E})^{1/k}} = \frac{C_{STD}}{(\ln\frac{n}{E_{STD}})^{1/k}} \qquad (3\text{-}14)$$

Thus, any combination of C_E and E satisfying Eq. (3-14) could serve as alternate equivalent specifications for the standard. However, the interrelationship is a function of k. The practical significance of this is that the effect of a trade-off between concentration level and exceedance rate would not be uniform for all monitoring sites since k is known to vary from site to site. Tables 3-1 and 3-2 show the trade-off for a base case of C_{STD} = 0.08 ppm, E_{STD} = 1. It is seen that the relationship is strongly influenced by the value of the shape factor k.

Table 3-1 Levels of Standards with Expected Exceedances per Year Equal to One Which are Equivalent to a Standard with a Level of 0.08 ppm but with Expected Exceedances per Year Greater than One

Expected Exceedances Per Year With Standard Level = 0.08 ppm		Equivalent Levels for Standards With Expected Exceedances/yr = 1, (ppm)				
	$k^{(1)}$ =	0.5	0.75	1	1.5	2
1		0.080	0.080	0.080	0.080	0.080
2		0.094	0.089	0.087	0.084	0.083
3		0.104	0.095	0.091	0.087	0.085
4		0.111	0.100	0.094	0.089	0.087
5		0.118	0.104	0.097	0.091	0.088
6		0.124	0.107	0.100	0.093	0.089
7		0.130	0.110	0.102	0.094	0.090
8		0.135	0.113	0.104	0.095	0.091
9		0.139	0.116	0.106	0.096	0.092
10		0.144	0.118	0.107	0.097	0.093
1		0.080	0.080	0.080	0.080	0.080
5		0.118	0.104	0.097	0.091	0.088
10		0.144	0.118	0.107	0.097	0.093
15		0.162	0.128	0.114	0.101	0.096
20		0.178	0.136	0.119	0.104	0.098
25		0.192	0.143	0.124	0.107	0.100
30		0.205	0.150	0.128	0.109	0.101
35		0.216	0.155	0.132	0.111	0.103
40		0.227	0.160	0.135	0.113	0.104
45		0.237	0.165	0.138	0.115	0.105
50		0.247	0.170	0.141	0.116	0.106

(1) k = constant in Weibull Distribution. Effective hours in year taken as 8760

Table 3-2 Levels of Standards with Expected Exceedances per Year Greater than One Equivalent to Standard with Level of 0.08 ppm and Expected Exceedance per Year Equal to One.

Expected	Equivalent Levels of Standards (ppm)				
Exceedances/yr	$k^{(1)}=$ 0.5	0.75	1	1.5	2
1	0.080	0.080	0.080	0.080	0.080
2	0.068	0.072	0.074	0.076	0.077
3	0.062	0.067	0.070	0.073	0.075
4	0.057	0.064	0.068	0.072	0.074
5	0.054	0.062	0.066	0.070	0.073
6	0.052	0.060	0.064	0.069	0.072
7	0.049	0.058	0.063	0.068	0.071
8	0.048	0.057	0.062	0.067	0.070
9	0.046	0.055	0.061	0.067	0.070
10	0.045	0.054	0.060	0.066	0.069
1	0.080	0.080	0.080	0.080	0.080
5	0.054	0.062	0.066	0.070	0.073
10	0.045	0.054	0.060	0.066	0.069
15	0.039	0.050	0.056	0.063	0.067
20	0.036	0.047	0.054	0.061	0.065
25	0.033	0.045	0.052	0.060	0.064
30	0.031	0.043	0.050	0.059	0.063
35	0.030	0.041	0.049	0.057	0.062
40	0.028	0.040	0.047	0.057	0.062
45	0.027	0.039	0.046	0.056	0.061
50	0.026	0.038	0.046	0.055	0.060

(1) k = constant in Weibull Distribution. Effective hours per year taken as 8760

Because of this strong dependence on k, ambient air quality standards with high values of E will on a U. S. wide basis provide less protection from high concentrations of pollutant than a standard with a low value of E. If k were uniform across the U. S. the exchange relationship Eq. (3-13) could be used without any degradation in degree of protection from extreme concentrations.

3.2.3 Statistical Behavior of the Extreme Values of the Hourly Average Photochemical Oxidant Concentrations Observed Over a Given Period

Short-term standards such as the photochemical oxidants standard are set for those pollutants for which short-term exposures are sufficient to cause undesirable effects. The standards are, therefore, designed to provide assurance that even the highest atmospheric concentrations observed in a one year period will not constitute a threat to the public health.

The short-term standards, therefore, focus on the highest concentrations likely to be observed in a given period, generally one year. It is of interest then, to study the behavior of the highest hourly average concentrations. In particular it is desired to know the distribution and probability density functions for the m th highest observation in n hourly average concentrations (n=8760 for one year). From these distributions one can calculate the median and other fractiles, the mean, the most probable value, and moments about the mean.

The distribution and density functions can be readily derived if the probability of occurrence of any hourly average concentration level is the same for each hour of the year. This is not the case for photochemical oxidants. Concentrations are strongly influenced by the time of day and time of year. Furthermore, the concentrations on successive hours are strongly correlated and peak hourly average concentrations on successive days tend to correlate as well. Put in another way, the underlying random or stochastic process giving rise to the fluctuations in oxidant concentrations is nonstationary and autocorrelated. While a start has been made[7] in accounting for nonstationarity and autocorrelation their full effect on the behavior of hourly average concentrations is still not well understood.

Although the results are only approximate it is instructive to study the statistics of the extreme values assuming the hourly average concentrations are independently and identically distributed. The starting point for the analysis will be the following basic relationship derived from probability theory[8].

$$P_m = \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m} \qquad (3\text{-}15)$$

where P_m is the probability that a certain event will occur m times in n independent trials, and where p is the probability of occurrence for a single trial. The collection of values, $P_0, P_1, \cdots$ etc., are referred to as the binomial distribution.

If we set G(C) = p, and take n as the number of hours in one year then P_m becomes the probability that the concentration C will be exceeded m times in one year.

We are also interested in the probability that a given concentration will be exceeded m or more times in a given year. If this probability is designated P_{m+} it follows that:

$$P_{m+} = 1 - \sum_{\nu=0}^{m-1} P_\nu \qquad (3\text{-}16)$$

Where P_ν is given by Eq. (3-15).

Note that in using Eqs. (3-15) and (3-16) the whole distribution need not necessarily be defined. The equations can be used for specific probabilities. For example, the probability for exceeding the concentration C_E is, based on the preceeding discussion, E/n. This quantity can then be substituted directly into Eq. (3-16) to calculate the probability of observing m actual exceedances in a year of a concentration level corresponding to an expected exceedance rate of E. Equation (3-16) can then be used to calculate probabilities for m or more actual exceedances. Since these calculations can be useful, the equations explicitly showing the expected exceedance rates are given here:

$$P_m = \frac{n!}{m!(n-m)!} \left(\frac{E}{n_E}\right)^m \left(1 - \frac{E}{n_E}\right)^{n-m} \qquad (3\text{-}17)$$

and

$$P_{m+} = 1 - \sum_{\nu=0}^{m-1} \frac{n!}{\nu!(n-\nu)!} \left(\frac{E}{n_E}\right)^{\nu} \left(1 - \frac{E}{n_E}\right)^{n-\nu} \qquad (3\text{-}18)$$

The term n_E is used in both equations since it specifically refers to the period of time over which the expected exceedance rate E occurs. The term n can be any number of hours and may be, for example, a multiple number of years.

Table 3-3 shows calculated P_m for a range of m and E for $n_E = n = 8760$. Table 3-4 shows calculated values for P_{m+} for both n_E and n equal to 8760 and 500. It is seen there is little difference between the two sets of probabilities in Table 3-4. In fact, to a good approximation for large n and n_E and small E and m Eq. (3-16) reduces to the following approximation relationship:

$$P_m \simeq \frac{\lambda E}{m!} e^{-\lambda E} \qquad (3\text{-}19)$$

Table 3-3 % Risk of m Exceedances/yr for E Expected Exceedances

$n_E = n = 8760$

m	E = 1	2	3	4	5	10
0	36.79	13.53	4.98	1.83	0.67	<0.01
1	36.79	27.07	14.93	7.32	3.37	0.05
2	18.40	27.07	22.41	14.65	8.42	0.23
3	6.13	18.05	22.42	19.55	14.04	0.75
4	1.53	9.03	16.82	19.55	17.56	1.89
5	0.31	3.61	10.09	15.65	17.57	3.78
6	0.05	1.20	5.05	10.44	14.65	6.31
7	0.01	0.34	2.16	5.97	10.47	9.03
8	<0.01	0.09	0.81	2.99	6.55	11.30
9		0.02	0.27	1.33	3.64	12.57
10		<0.01	0.08	0.53	1.82	12.58
11			0.02	0.19	0.83	11.45
12			0.01	0.06	0.35	9.55
13			<0.01	0.02	0.13	7.36
14				0.01	0.05	5.26
15				<0.01	0.02	3.51
16					<0.01	2.20
17						1.29
18						0.72
19						0.38
20						0.19
21						0.09
22						0.04
23						0.02
24						0.01
25						<0.01

Table 3-4 % Risk of m or More Exceedances/yr of a Given Standard Level for E Expected Exceedances

E	m = 1	2	3	4	5	10
$n_E = n = 8760$						
1	63.2	26.4	8.0	1.9	0.4	< 0.1
2	86.5	59.4	32.3	14.3	5.3	< 0.1
3	95.0	80.1	57.7	35.3	18.5	< 0.1
4	98.2	90.8	76.2	56.7	37.1	0.8
5	99.3	96.0	87.5	73.5	56.0	3.2
10	> 99.9	99.9	99.7	99.0	97.0	54.2
$n_E = n = 500$						
1	63.2	26.4	8.0	1.9	0.4	< 0.1
2	86.5	59.4	32.3	14.3	5.2	< 0.1
3	95.1	80.2	57.8	35.3	18.4	0.1
4	98.2	90.9	76.3	56.7	37.1	0.8
5	99.3	96.0	87.7	73.6	56.0	3.1
10	> 99.9	> 99.9	99.7	99.0	97.2	54.3

where:

$$\lambda = \frac{n}{n_E} \tag{3-20}$$

The derivation of the expression (3-18) is given in Note 1 of Appendix A.

Coming back to Eq. (3-15), if P_m is obtained for $m = 0$ the following simple expression results:

$$P_0 = (1 - p)^n \tag{3-21}$$

or

$$P_0 = (1 - G(C))^n \tag{3-22}$$

where $G(C)$ is the distribution function for the hourly average concentrations. The P_0 of Eq. (3-22) is the probability that the concentration C will not be exceeded in n hours. It is, therefore, also the distribution function for the highest observed concentration in n observations. That is, if we define $F(C_{(m)})$ as the distribution function for the mth highest value in n hours (observations) then:

$$F(C_{(1)}) = P_0 \tag{3-23}$$

To obtain $F(C_{(m)})$ for any m it is noted that all possible cases in which the mth highest value does not exceed a given concentration level are contained in those cases in which there are 0, 1, 2, 3, ····, or m-1 exceedances. The probabilities for these cases are P_0, P_1, ⋯, and P_{m-1}. By the additive law of probability then:

$$F(C_{(m)}) = \sum_{\nu=0}^{m-1} P_\nu \tag{3-24}$$

also from Eq. (3-16)

$$F(C_{(m)}) = 1 - P_{m+} \tag{3-25}$$

Eq. (3-24) may be written more explicitly as:

$$F(C_{(m)}) = \sum_{\nu=0}^{m-1} \frac{n!}{\nu!(n-\nu)!} p^\nu (1-p)^{n-\nu} \tag{3-26}$$

To obtain the corresponding probability density function Eq. (3-26) is differentiated with respect to C to yield:

$$f(C_{(m)}) = \frac{n!}{(m-1)(n-m)!} (1-p)^{n-m} p^{m-1} (- \frac{dp}{dc}) \tag{3-27}$$

The median of the distribution for the mth highest concentration is obtained by substituting G(C) for p in Eq. (3-26) and finding the value of C which yields $F(C_{(m)}) = 0.5$. The mean is the expected value of $C_{(m)}$ and by definition is:

$$\overline{C}_{(m)} = \int_0^\infty C\ f(C_{(m)})dC \tag{3-28}$$

The most probable value, or mode, of the mth highest value is found by setting derivative of $f(C_{(m)})$ equal to zero and determining the value of $C_{(m)}$ which satisfies the expression.

For the Weibull distribution the distribution of the highest value in n hours is found by substituting Eq. (3-10) into Eq. (3-24) (or in this case, Eq. (3-21), which yields:

$$F(C_{(1)}) = \left(1-e^{-(\ln\frac{n_E}{E})\,(C/C_E)^k}\right)^n \qquad (3\text{-}29)$$

The probability density function is obtained from Eq. (3-25):

$$f(C_{(1)}) = \frac{nk}{C_E}\left(\ln\frac{n_E}{E}\right)(C/C_E)^{k-1}\left(1-e^{-(\ln\frac{n_E}{E})(C/C_E)^k}\right)^{n-1} \cdot$$

$$e^{-(\ln\frac{n_E}{E})\,(C/C_E)^k} \qquad (3\text{-}30)$$

Figure 3-2 shows the distribution of the annual maximum value for a Weibull distribution of hourly average concentrations, Eq. (3-29), for several values of the shape factor k. Each curve has the same value for C_E, E, and n and, therefore, goes through the same point at C = 1. Note, the effect of increasing k is to narrow the range of the distribution. Because of the term C/C_E in Eq. (3-29), and the use of $C_E = 1$, the curves in Fig. 3-2 are easily transformed to any other value of C_E by multiplying the concentration axis values by the new C_E. In effect the concentration axis in Fig. 3-2 could be labeled relative concentration. It is relative to the value of C_E as plotted.

Figure 3-2 Effect of Weibull Distribution Shape Factor on Distribution of Highest Value

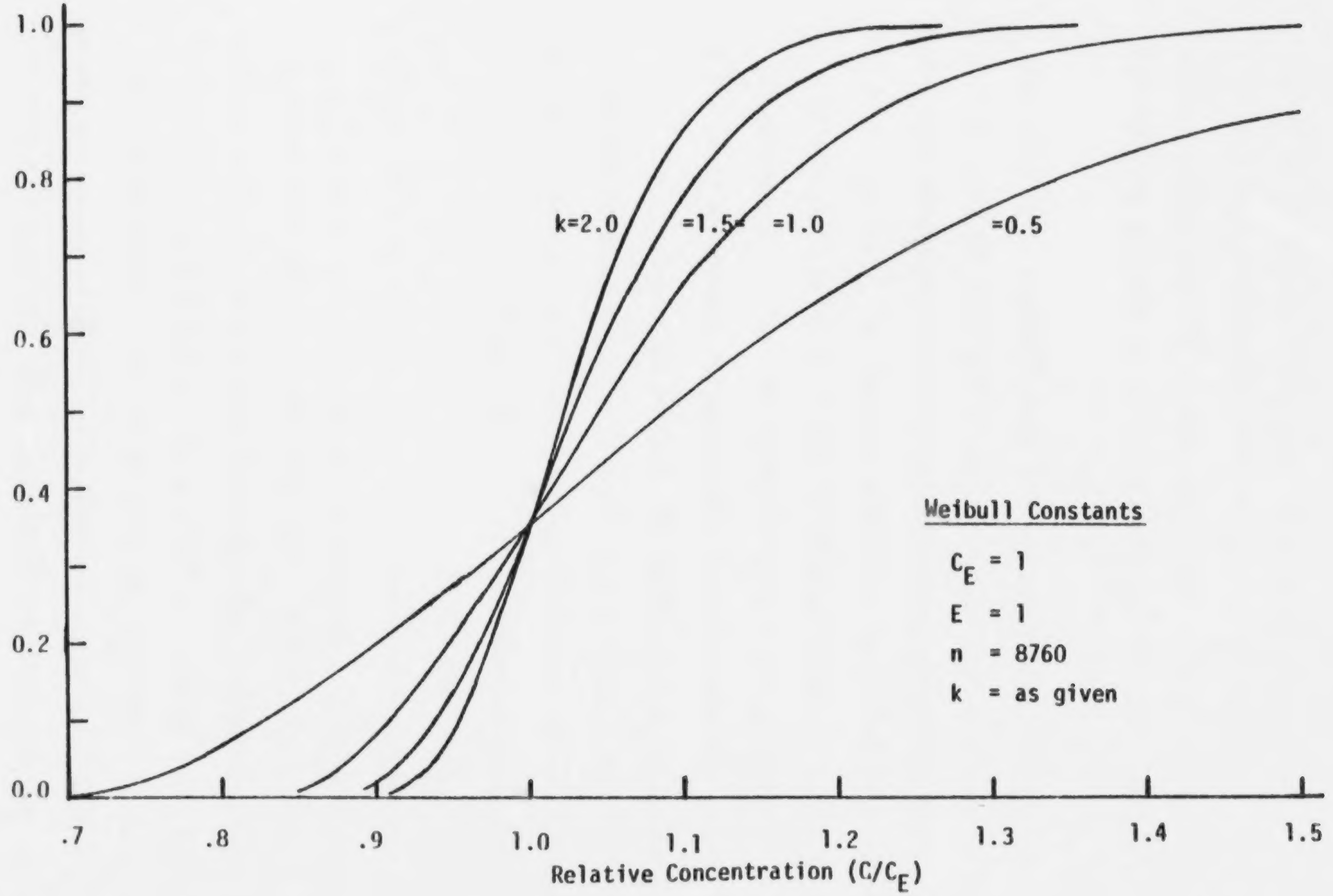


Figure 3-3 shows the corresponding density functions Eq. (3-30). It is seen that the most probable value of the maximum concentration is very close to C = 1. In fact, for k = 1, the most probable value of C is C_E for E = 1 (designated as C_1). For k less than 1, the most probable value is slightly below C_1 and for k greater than 1 it is slightly greater than C_1.

The cumulative distributions for the first, second, ····, fifth and tenth highest hourly average concentrations in one year are shown in Fig. 3-4 and the corresponding probability density functions in Fig. 3-5 for a Weibull distribution of hourly averages with k = 1 and $C_E = 1$, E = 1. As indicated above the distributions for any other value of C_E can be obtained by multiplying the x-axis values by the new C_E. The maxima in Fig. 3-5 fall on the concentrations whose expected exceedance rates are m. The values of these concentrations are shown in Table 3-5.

It is shown in Appendix A that the most probable value of the mth highest hourly average concentration in one year is the concentration, C_E, whose expected exceedance rate, E, is m for an exponential distribution (Weibull distribution with k = 1).

Table 3-6 shows the locations of the most probable value of the mth highest concentration as a function of the Weibull shape factor for $n = n_E = 500$ and $n = n_E = 8760$. The lower value of n has application in cases of nonindependence of hours and will be discussed in latter sections. Here it is of value because the position of the most probable value is more sensitive to k the lower the value of n. For k values above 0.75 the most probable value of the mth highest concentration can be considered equal to the concentration with an expected exceedance rate of m with no significant error.

Figure 3-3 Probability Density Function of Annual Highest Concentration for Different Values of Weibull Distribution Shape Factor

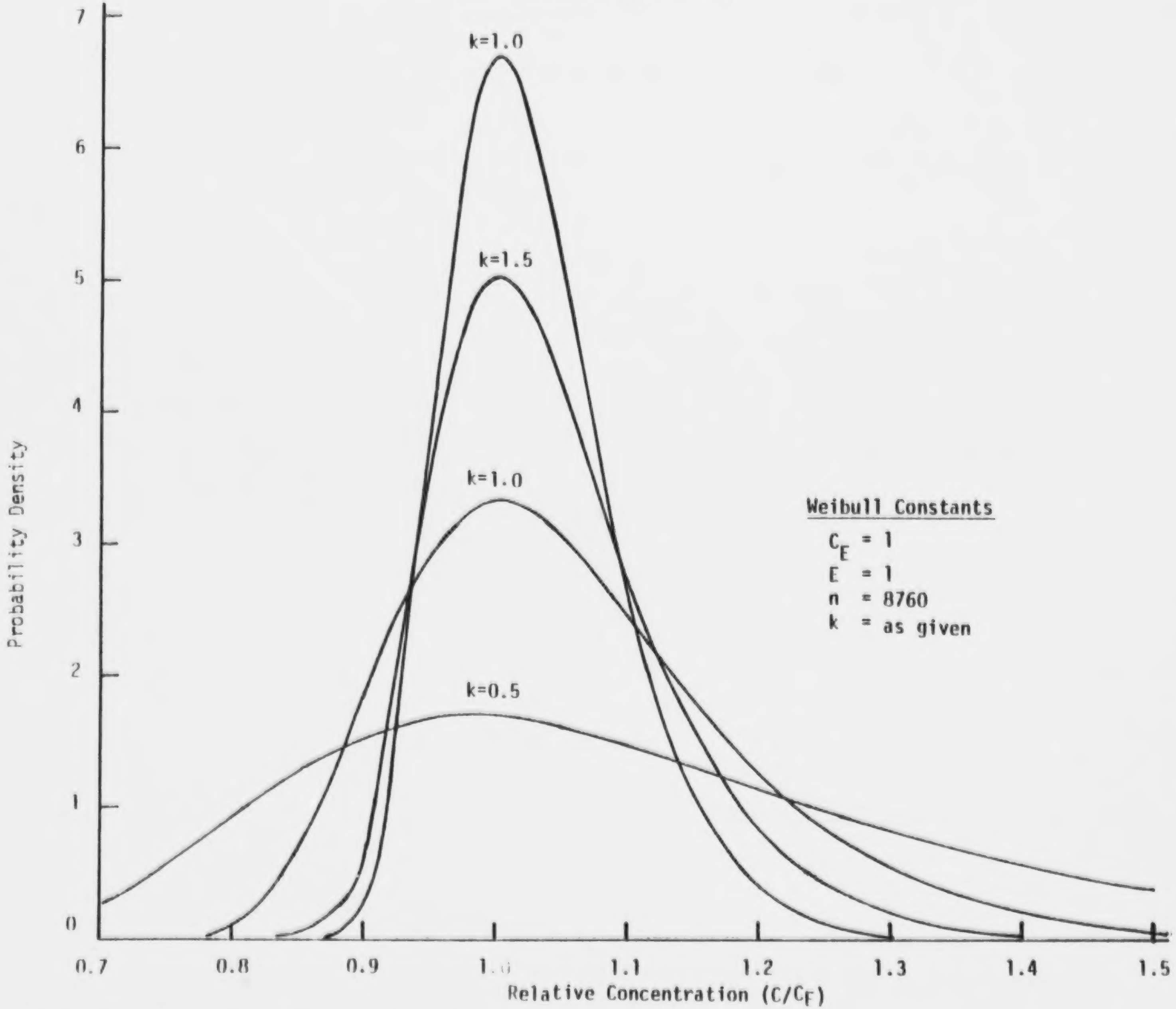


Figure 3-4 Distribution of Annual Extremal Values For a Weibull Distribution of Hourly Concentrations Assuming Independent Hours

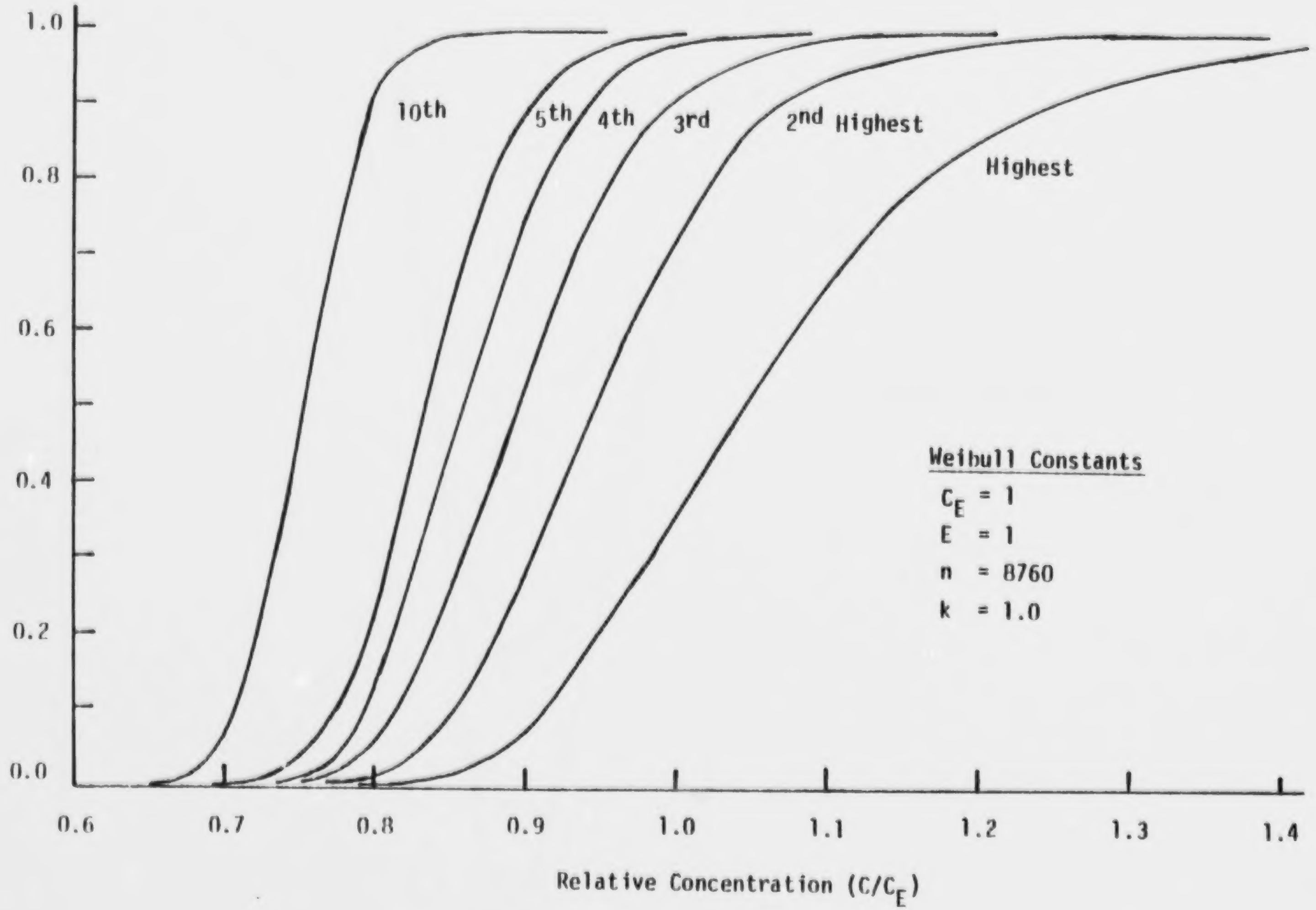


Figure 3-5 Probability Density Functions for Annual Extremal Values for a Weibull Distribution of Hourly Average Concentrations (Assumes Independence of Hours)

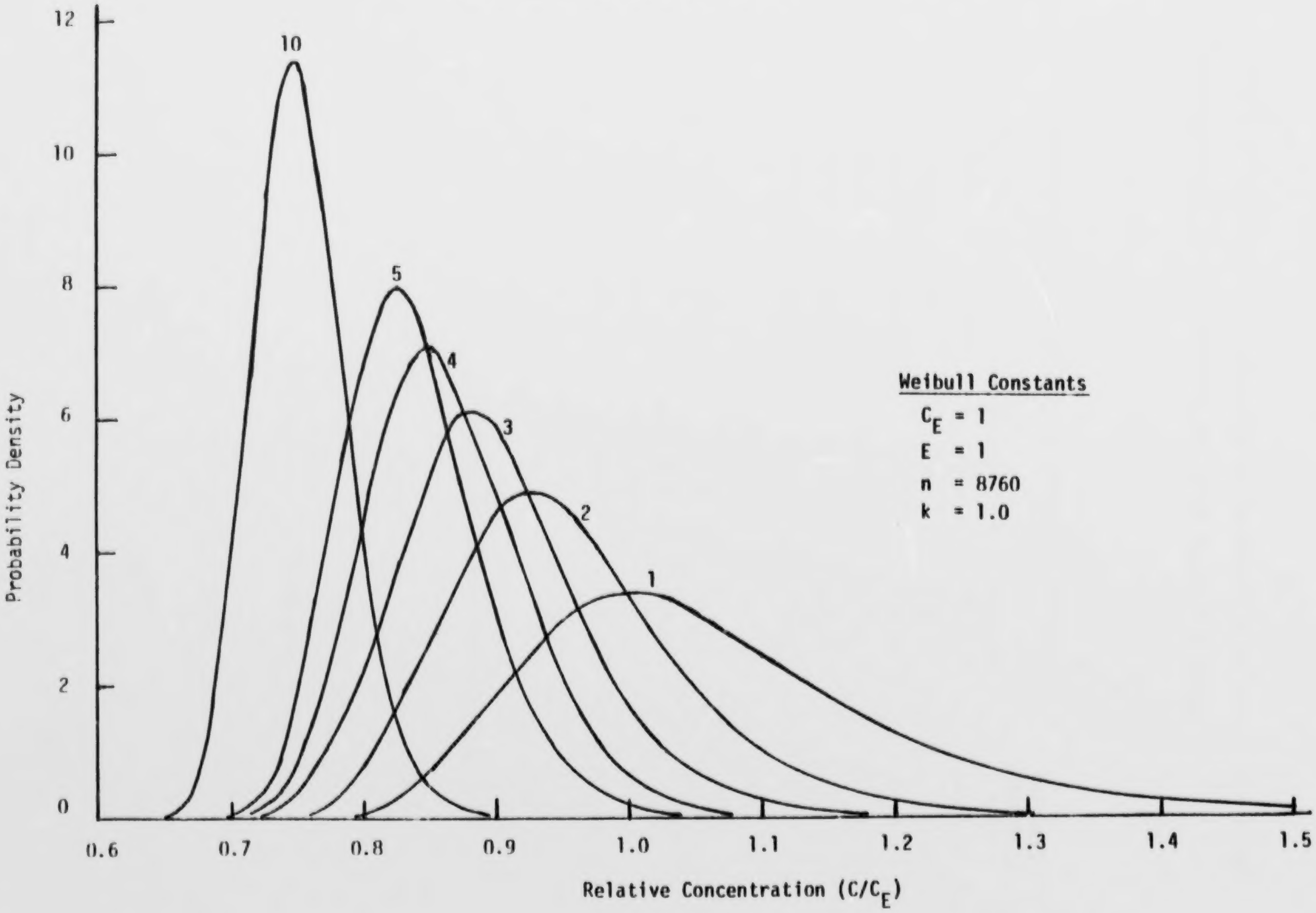


Table 3-5 Relative Positions of Concentrations Corresponding to Various Expected Exceedance Rates for Different Weibull Distributions(1)

Expected Exceedance Rate E	Relative Concentration for Given Exceedance Rate C_E/C_1 k = 0.5	=1.0	=1.5	=2.0
1	1.000	1.000	1.000	1.000
2	0.853	0.924	0.948	0.961
3	0.773	0.879	0.918	0.938
4	0.718	0.847	0.895	0.920
5	0.677	0.823	0.878	0.907
10	0.557	0.746	0.823	0.864
25	0.417	0.645	0.747	0.803
50	0.324	0.569	0.687	0.754
100	0.243	0.493	0.624	0.702

(1)Weibull Distribution Constants

C_E = 1

E = 1

n = 8760

k = as given

Table 3-6 Location of the Most Probable Value of mth Highest Concentration as a Function of Number of Hours, n and Weibull Shape Factor, k

Rank m	Concentration Level of mth Highest Value Relative to Concentration with Expected Exceedance Rate of m k = 0.5	= 1.0	= 1.5	= 2.0
For n and n_E = 8760				
1	0.977	1.000	1.003	1.003
For n and n_E = 500				
1	0.951	1.000	1.006	1.006
2	0.969	1.000	1.004	1.004
3	0.975	1.000	1.003	1.004
4	0.975	1.000	1.003	1.003
5	0.981	1.000	1.003	1.003
10	0.987	1.000	1.001	1.001

The relative location of the expected or average value of the mth highest concentration is shown in Table 3-7. The expected value of the mth highest concentration is close to the concentration whose expected exceedance rate is m. However, it is always distinctly higher because the distributions are skewed to the high side of the mode.

Table 3-7 Location of Expected Value of the mth Highest Concentration for an Exponential Distribution(1) Relative to the Concentration with an Expected Exceedance Rate m for Two Levels of n

Rank	Relative Expected Value of mth Highest Concentration	
m	$n = n_E = 500$	$n = n_E = 8760$
1	1.093	1.050
2	1.050	1.026
3	1.035	1.021
4	1.028	1.016
5	1.023	1.013
10	1.014	1.007

(1) Weibull Distribution with k = 1

3.3 Nonindependence of Ambient Concentrations of Pollutants

The previous section in dealing with the behavior of time averaged concentrations of pollutants assumed that there existed no correlation between the concentrations of a pollutant observed in one hour and that observed in another hour. It further assumed that the underlying stochastic processes giving rise to observed concentrations were constant in time, or stationary. This is not the case for most air pollutants. In the case of ozone the concentration in one hour tends to be dependent on concentations in previous hours. Furthermore, the peak concentrations observed in one day tend to correlate with the peak concentrations observed the next day. This correlation can persist over several days. There is a strong time dependence. Ozone concentations over a twenty-four hour period show fairly well defined patterns in urban areas which relate to the degree of sunlight and vehicle use. There can also be a dependence on day of the week. And there is a strong dependence on time of year, with the highest concentrations occurring during the warmer months of the year in urban areas.

The extent to which nonstationarity in the processes leading to observed ambient concentrations of pollutants and autocorrelation of neighboring concentrations must be accounted for depends upon the form of the standard and the information desired. In the case of the proposed ozone standard there is no problem in either determining compliance or degree of departure from compliance. In determining compliance either the average number of exceedances per year of the standard or C_1, the concentration of corresponding to an expected exceedance rate of one, needs to be determined. In deciding the degree of departure from the standard, C_1 is needed. The average exceedance rate can be obtained by direct measure or from the distribution of hourly average concentrations. The concentration C_1 can be obtained from the distribution. The question of independence does not enter.

If the standard was based on the expected annual maximum hourly average concentration or the expected annual second highest hourly average concentration, the question of dependence would have to be dealt with. As seen in Section 3.2.3 in order to estimate the expected maximum or expected second high concentration it is necessary to know the distributions for the maximum and second high concentrations respectively. In order to derive the exact distributions it would be necessary to fully specify the time dependence and autocorrelation of the hourly average concentrations. The problems of dependence also becomes important in estimating the risk associated with alternative standard levels (Section 4).

Accounting for dependence in an exact manner would require the accurate modeling of the nonstationarity and autocorrelation and would seem to present formidable analytical difficulties. The study performed under this contract was confined, therefore, to approximate methods of dealing with the problem. The simplest approximation, of course, is to assume dependence and apply the expressions developed in Section 3.2.3. An EPA study[7] of the problem of dependence associated with the daily maximum hourly average ozone concentration concluded that the major problem in assumption of independence was nonstationarity. If this is the case then the approximation can be used if the highest concentrations in a one year period occur in a subset of the hours in a year for which the underlying stochastic processes are relatively stationary. For ozone a possible subset would be the midday hours during the summer months, for example, 11 A.M. to 4 P.M. June 15 through September 15. This period would contain close to 500 hours. (Aerometric data would need to be examined to determine what size the subset needs to be to have reasonable assurance of capturing the maximum concentrations.) Over the June to September period the day to day generating processes can be assumed reasonably stationary. The degree of correlation from one day to the next would be low to moderate. Within the four hour segment of any day the process could be considered relatively stationary with moderate to strong correlation. If we assume the strong correlation within a day would not have a large effect on the approximation, then the expressions in Section 3.2.3 can be used if 500 hours is substituted in the place of 8760 hours.

In making the above mentioned substitution it must be remembered that the distribution of hourly averages, G(C), now applies to the 500 hour segment instead of the full 8760 hours in one year. Thus, it will be necessary to test whether the Weibull distribution is still applicable, and if it is, whether there is a change in the Weibull constants, δ and k. (See Eq. (3-6)). Since δ is largely determined by the ambient air standard (Eq. (3-9)) the problem is basically that of determining the effect on the shape factor k.

The effect on k of using a subset of the 8760 hours in one year was examined using two very approximate approaches. In the first approach, it was assumed that the 500 highest ozone concentrations of the 8760 hours of measurements contained in the data set all fell within the summer subset (approximately the five midday hours from June 15 to September 15). This is, of course, a very approximate assumption.

If the n highest hours are contained within n_o hours and a Weibull distribution fits the n_o hours of data, we can write:

$$g(C) = e^{-(\ln n_o)\left(\frac{C}{C_1}\right)^k} \tag{3-31}$$

where C_1 is the concentration corresponding to an expected exceedance rate of 1 in n_o hours. We can write:

$$\frac{E_j}{n_o} = e^{-(\ln n_o)\left(\frac{C_j}{C_1}\right)} \tag{3-32}$$

and further that:

$$\frac{E_j}{n} = \frac{n_o}{n} e^{-(\ln n_o)\left(\frac{C_j}{C_1}\right)^k} \tag{3-33}$$

Thus, if the n highest values were replotted using n as the devisor to calculate a frequency factor for each data point, the resulting expression (Eq. (3-33)) can be thought of as a three parameter Weibull, where n_o/n is the third parameter. Thus the effect of narrowing in on a segment of hours tends to be to introduce another constant into the original Weibull distribution. The third parameter tends to be a displacement factor since Eq. (3-33) can be rewritten to give:

$$\frac{E_j}{n} = e^{-(\ln n_o)\left[\left(\frac{C_j}{C_1}\right)^k - \frac{\ln n_o/n}{\ln n_o}\right]} \tag{3-34}$$

The distribution of the maximum hourly average concentration for the distribution given by Eq. (3-33) is:

$$P_o = (1 - \frac{n_o}{n} e^{-(\ln n_o)\,(\frac{C}{C_1})^k})^n \tag{3-35}$$

It turns out that P_o calculated by Eq. (3-35) is very close to that which would be calculated using Eq. (3-31):

$$P_o = (1 - e^{-(\ln n_o)\,(\frac{C}{C_1})^k})^{n_o} \tag{3-36}$$

that is:

$$(1 - p)^{n_o} \simeq (1 - \frac{n_o}{n} p)^n \tag{3-37}$$

The result can also be verified by direct calculation as shown in Table 3-8. It is seen from Table 3-8 that for $n^o = 8760$ and $n = 500$ the distribution of the maximum value is almost identical for either $k = 1$ or 0.5. Even for n as low as 10 the approximation holds rather well. Thus, if many of the highest values were included in the smaller time segment, there would be no effective change in k, and the new distribution would be most effectively described by a three parameter Weibull. The more interesting conclusion is that the nonstationarity leading to the segmentation would have little effect on the distribution of the highest hourly average concentration. That is, Eq. (3-36) could be used to calculate the distribution. While Eq. (3-36) could not be said to hold down to $n = 500$ for $n_o = 8760$, it may apply

Table 3-8 Comparison of Distributions of Annual Maximum Hourly Average Concentration as Determined by Two Related Hourly Distributions

Relative Concentration (C/C_1)	Weibull Distribution(1)		Modified Weibull Distribution(2)		
	$n_o=8760$ $k=1$	$n_o=8760$ $k=0.5$	$n=500$ $k=1$	$n=500$ $k=0.5$	$n=10$ $k=0.5$
0.7	0	0.012	0	0.012	0.003
0.8	0.002	0.074	0.002	0.073	0.049
0.9	0.084	0.203	0.083	0.203	0.176
1.0	0.368	0.368	0.368	0.368	0.349
1.1	0.668	0.526	0.668	0.526	0.515
1.2	0.850	0.657	0.850	0.657	0.651
1.3	0.936	0.756	0.936	0.756	0.753
1.4	0.974	0.827	0.974	0.827	0.826
1.5	0.989	0.878	0.989	0.878	0.877
1.6	0.996	0.914	0.996	0.914	0.913
1.7	0.998	0.939	0.998	0.939	0.938
1.8	0.999	0.956	0.999	0.956	0.956
1.9	1.000	0.968	1.000	0.968	0.968
2.0	1.000	0.977	1.000	0.977	0.977

(1) $G(X) = \exp\left(-(\ln n_o)(C/C_1)^k\right)$,

(2) $G(X) = (n/n_o)\exp\left(-(\ln n_o)(C/C_1)^k\right)$, $n_o = 8760$

fairly well to n = 10 (the ten highest numbers). Thus, this result is suggestive that Eq. (3-36) gives a reasonable approximation for the distribution of the highest ozone hourly average concentrations.

If a summer segment of selected times of the day is plotted, it probably will not contain the 500 highest hours and therefore, will have a higher k than indicated by Eq. (3-33). To investigate this point further the second approach studied the question as to what value of k is required for an hourly average concentration distribution for n = 500 to give an approximately equal maximum value distribution to that obtained with n = 8760. Base case maximum value distributions were derived for n_0 = 8760 and k = 0.5, 1.0, and 2.0. In each case maximum value distributions for n_0 = 500, using Eq. (3-36), were calculated for a range of k values to see which k gave the best fit to the base case. In each case the best fit k value for n = 500 was larger than the k for n = 8760 by a factor of about 1.4. The results for base case k = 1.0 are shown in Table 3-9.

It was also of interest to repeat the above calculations for the distribution of the fifth highest concentration. The results are shown in Table 3-10. They show that the modified Weibull continues to provide a good approximation of the distribution of the fifth high. In the case of the unmodified Weibull, the factor by which k must be increased to bring the distributions into alignment increases from 1.4 to somewhat less than 1.6.

In a parallel investigation PEDCo Environmental, Inc.[8] examined aerometric ozone data from a number of monitoring sites for various subsets of the hours in a year and compared the k values for Weibull distributions fitted to the subsets with those obtained for the full year of data. Table 3-11 shows the k values obtained for Weibull distributions fitted to data from several locations for the complete set of 8760 hours and two subsets,

Table 3-9 Effective Increase in Weibull Constant k Needed to Give Same Distribution of Maximum Value(1) as n is Decreased - Base Case: k=1.0

Relative Concentration C/C_1	Base Case Distribution of Maximum n=8760 k=1.0	Distribution of Maximum for n = 500 k=1.0	1.3	1.4(2)	1.5	2.0
0.7	0.000	0.002	0.000	0.000	0.000	0.000
0.8	0.002	0.031	0.008	0.005	0.003	0.000
0.9	0.084	0.155	0.109	0.095	0.083	0.038
1.0	0.368	0.368	0.368	0.368	0.368	0.368
1.1	0.668	0.584	0.644	0.662	0.681	0.762
1.2	0.850	0.749	0.827	0.849	0.868	0.937
1.3	0.936	0.856	0.923	0.939	0.951	0.986
1.4	0.974	0.920	0.967	0.977	0.983	0.997
1.5	0.989	0.956	0.987	0.991	0.995	
1.6	0.996	0.976	0.995	0.997	0.998	
1.7	0.998	0.987	0.998	0.999	0.999	
1.8	0.999	0.993	0.100	1.000	1.000	
1.9	1.000	0.996				
2.0		0.998				

(1) Distribution of Maximum Value = $[1 - \exp(-(\ln n)(C/C_1)^k)]^n$

(2) Closest fit k = 1.4; $k_1/k_2 = 1.4/1.0 = 1.4$

Table 3-10 Effective Increase in Weibull Constant k Needed to Give Same Distribution of Maximum Value(1) as n is Decreased - Base Case: k=0.5

Relative Concentration C/C_1	Base Case Distribution of Maximum n=8760 k=0.5	Distribution of Maximum for n = 500					
		k=0.5	=0.7(2)	=0.73	=0.75	=0.8	=1.0
0.7	0.012	0.063	0.019	0.015	0.013	0.009	0.002
0.8	0.074	0.145	0.085	0.078	0.073	0.063	0.031
0.9	0.203	0.252	0.211	0.205	0.201	0.191	0.155
1.0	0.368	0.368	0.368	0.368	0.368	0.368	0.368
1.1	0.526	0.478	0.521	0.528	0.532	0.543	0.584
1.2	0.657	0.575	0.651	0.662	0.669	0.686	0.749
1.3	0.756	0.658	0.751	0.764	0.772	0.791	0.856
1.4	0.827	0.726	0.825	0.838	0.845	0.864	0.920
1.5	0.878	0.781	0.878	0.889	0.896	0.912	0.956
1.6	0.914	0.825	0.915	0.924	0.930	0.943	0.976
1.7	0.939	0.860	0.941	0.949	0.953	0.963	0.987
1.8	0.956	0.887	0.959	0.965	0.969	0.976	0.993
1.9	0.968	0.909	0.971	0.976	0.979	0.985	0.996
2.0	0.977	0.927	0.980	0.983	0.986	0.990	0.998

(1) See Table 3 for defining equation.

(2) Closest fit k = 0.7; $\frac{k_1}{k_2} = \frac{0.7}{0.5} = 1.4$

Table 3-11 Change in Weibull Distribution Shape Factors with Changing Time Segment in Which Hourly Average Concentrations are Collected(1)

Site	Year	Shape Factor, k Full Year(2)	May-Sept.(3) 11 AM-6 PM	July-Aug.(4) 11 AM-6 PM
Kansas City, Kansas	1975	1.24	--	4.34(5)
Des Moines, Iowa	1975	1.92	2.04	2.40
Louisville, Kentucky	1974	0.80	1.66	2.02
Memphis, Tennessee	1974	1.32	2.28	2.34
Mamaroneck, New York	1975	0.80	1.31	1.57
Racine, Wisconsin	1974	1.49	--	1.99(5)

(1) Reference 8

(2) 8760 hours.

(3) 1071 hours.

(4) 434 hours.

(5) Poor Weibull fit.

one containing 1071 hours and the other 434 hours. They found that good fits of the Weibull distribution could be made to the subsets for most data sets. There was a marked increase in the k value. They recommend a subset of 1071 (11 AM to 6 PM, May-September) hours having a good chance of capturing the highest annual hourly average ozone concentration. Ignoring the two cases in the table in which a poor fit was obtained the k increased for the 1071 hour subset by an average factor of 1.6 (range 1.06-2.08) and for the 434 hour subset by an average factor of 1.9 (range 1.25-2.53). These values are larger than suggested by the above approximate analysis.

In summary, it would appear that in the case of ambient hourly average concentrations of ozone the distributions and related statistics for the annual highest concentrations can be obtained by applying the assumption of independence to a subset of hours consisting of the 1071 hours in the period 11 AM to 6 PM, May through June, and by applying the k values corresponding to the subset.

3.4 Comparison of Statistical and Deterministic Standards

In changing ambient air quality standards from their present deterministic form to statistical forms the question arises as to how the standards compare in the air quality provided. A difficulty in making such comparisons is the uncertainty in the degree to which a deterministic standard would be enforced. As discussed earlier, strict enforcement of the present photochemical oxidant standard logically means that the overall air quality for an Air Quality Control Region (AQCR) must be brought the level that the 0.08 ppm hourly average concentration level is never exceeded. However, it seems unlikely that this would be done. For example, if conditions were such that the standard level was exceeded two or more times in one year (a violation of the standard) on an average of once every one hundred years it is questionable as to whether any effort would be made to further reduce precursor emissions to assure that two or more exceedances in a single year would never occur. The occurrence very likely would be downplayed as a freak meteorological condition untypical of the area or an instrument error and, therefore, not worthy of further control effort. It is difficult to say at what frequency of violation of the standard control officials would feel obliged to act. For some an acceptable average period of reoccurrence might be as low as ten or fifteen years. This difficulty stems directly from the nature of deterministic standards which ignore the statistical nature of ambient pollutant concentrations and needs to be kept in mind when comparing deterministic and statistical forms.

3.4.1 Comparison Based on Return Period of Violation

One way in which statistical standards can be compared with deterministic standards set at the same concentration level is through use of the return period. The return period is the average period of time separating the repitition of an event whose occurrence is under the influence of statistical processes. The return period is reciprocal of the relative frequency or probability of occurrence of the event. For example, if theevent is the occurrence of two exceedances of the standard level in one year and the relative frequency or probability of occurrence is 0.1, then the return period is ten years. That is, the event is expected to occur on an average of once every ten years.

A statistical form of the standard fully takes into account the statistical behavior of pollutant concentrations. Thus, in setting a statistical standard at a given concentration level it is acknowledged that there are nonzero probabilities that the standard level will be exceeded 0, 1, 2, or more times in the same year. These probabilities can be estimated. For example, Eq. (3-15) can be used as a basis for estimating the probability of m exceedances in one year of the level of a statistical standard and Eq. (3-16) can be used to estimate the probabilities of m or more exceedances in a one year period. The reciprocals of these probabilities are the return periods. In particular, in comparing alternative statistical forms with the present photochemical oxidant standard what is desired is the probability and corresponding return period associated with two or more exceedances of the standard level in a given year, since such an occurrence constitutes a violation of the deterministic standard.

In determining the probabilities it is most convenient to calculate them as functions of the expected exceedance rate. In this case Eqs. (3-17) and (3-18) are used. Figures 3-6 and 3-7 show the probabilities of exceedances and of m or more exceedances in one year as a function of the expected exceedance rate for a range of values of m. The curves are independent of the standard level and the distriubtion of hourly average concentrations. They are also nearly independent of the number of effective hours in one year. That is, the curves for 8760 hours are essentially identical to those for 1071 hours (see Section 3.2.3). It will be recalled from the previous section that the use of approximately 1000 effective hours in a year tends to compensate for problems of nonstationarity in hourly average ozone concentrations.

Figures 3-6 and 3-7 illustrate the problem of the present deterministic standard. Even at a low expected exceedance rate (e.g., E = 0.1) there is a nonzero probability of 2, 3, or more exceedances. To drive these probabilities to zero it is necessary to drive the probability of one exceedance (which is seemingly allowed by the present deterministic standard) to zero and the probability of no exceedances to one. Table 3-12 further illustrates this point with selected data points from the curves in Fig. 3-6. Each row represents the probabilities for 0, 1, 2, 3, 4, 5 actual exceedances in a year of a statistical standard level whose allowed expected exceedance rate is given in the left hand column. The proposed form of the ozone standard allows an expected exceedance rate of one. It is seen from Table 3-12 that there is a 37% chance in any year of no violation, a 37% chance of one violation and that in an AQCR just meeting the proposed standard there is a 37% chance a single violation will occur in a given year and an 18% chance there will be two violations. By going to an expected exceedance rate of 0.1, the probability of two exceednaces can be reduced to 0.5 % or a return period of 200 years. The probability of one exceedance then, is forced to 9% or a return period of 11 years.

Figure 3-6 Probability of a Given Observed Number of Exceedances as a Function of the Expected Exceedance Rate

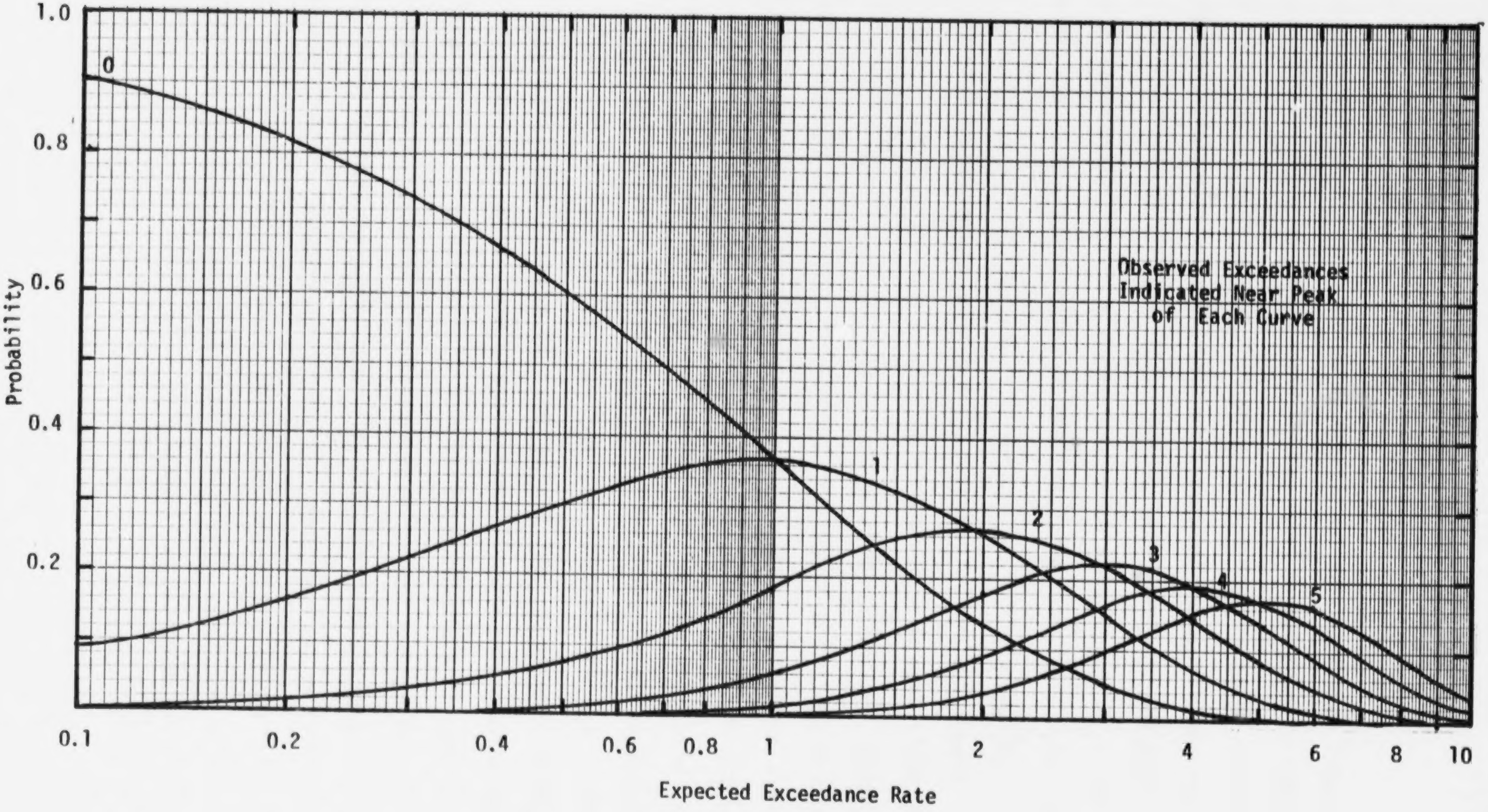


Figure 3-7 Probability of a Actual Exceedances Equal to or Above a Given Number as a Function of the Expected Exceedance Rate

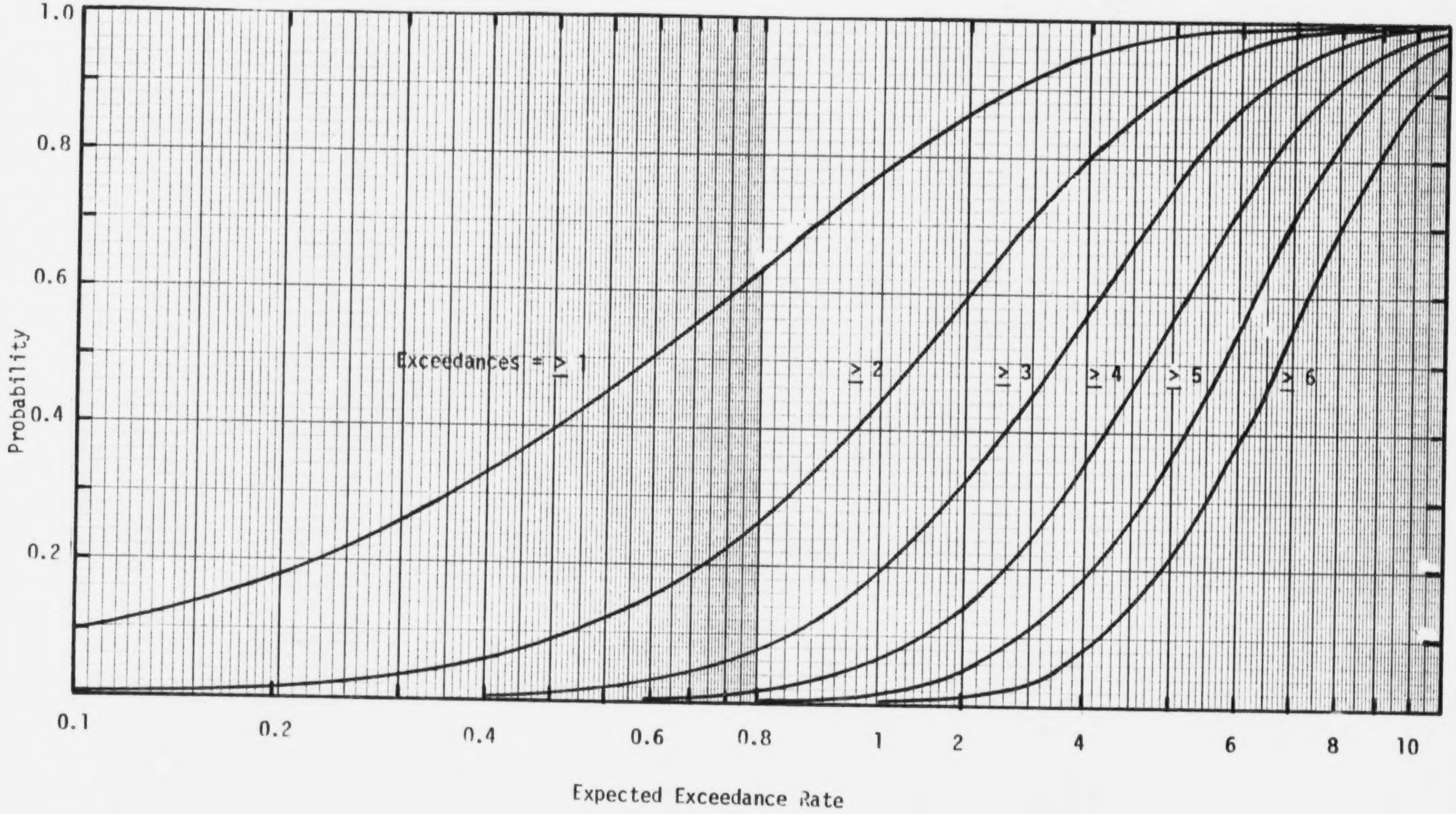


Table 3-12 Probability of a Given Observed Number of Exceedances for Different Expected Exceedance Rates

Expected Exceedance Rate	Actual Exceedances =	Probability					
		0	1	2	3	4	5
0.1		0.905	0.090	0.005	<0.001	<0.001	<0.001
0.2		0.819	0.164	0.016	0.001	<0.001	<0.001
0.5		0.606	0.303	0.076	0.013	0.002	<0.001
1		0.368	0.368	0.184	0.061	0.015	0.003
2		0.135	0.271	0.271	0.181	0.090	0.036
5		0.007	0.033	0.084	0.140	0.176	0.176
10		<0.000	<0.00	0.002	0.007	0.019	0.037

Table 3-13 shows the probability of two or more exceedances of the standard level and corresponding return period for standards with different allowed expected exceedance rates. It is seen that for the statistical form of the standard at an expected exceedance rate of one the corresponding deterministic form (not more than one exceedance of the standard level) with the same concentration level will be violated on an average of about once every four years. If the expected exceedance rate of the statistical standard was set at 0.1 the present deterministic standard would be violated on the average of about once every 200 years. Note also that as the expected exceedance rate is increased above one the return period rapidly falls to the point that the deterministic standard is being violated every year. Thus, by adjusting the exceedance rate the statistical standard can be made to approach as close as is desired the deterministic standard in stringency while maintaining the same concentration level. Of course, and as discussed in Section 3.2.2, this can also be achieved by maintaining the exceedance rate at one and lowering the concentration level of the statistical standard.

Table 3-13 also shows how statistical standards based on the expected annual maximum hourly average concentration or the expected annual second highest concentration compare with the present deterministic standard. Probabilities for these forms are calculated by the first calculating the expected maximum and expected second highest concentrations when the concentration, C_1, corresponding to an expected exceedance rate of one is equal to one (as done in Table 3-6) and then using Eq. (3-14) to calculate the expected exceedance rate at the expected highest and second highest concentrations. Given the expected exceedance rate the probability of two or more

Table 3-13 Comparison of Statistical Forms with a Deterministic Standard Which Allows No More than One Exceedance per Year

Expected Exceedance Rate	Probability of Two or More Exceedances	Return Period (years)
0.1	0.005	200.0
0.5	0.090	11.1
1.0	0.264	3.8
2.0	0.594	1.7
3.0	0.801	1.2
4.0	0.909	1.1
5.0	0.960	1.0
10.0	1.000	1.0
Expected Max	0.127	7.9
Expected Second High	0.461	2.2

exceedances in a given year is calculated from Eq. (3-18). Following the discussion in the previous section, 1071 effective hours and a range of Weibull shape factor, k, of 1.36 to 2.54 were used in the first part of the calculations. When this was done the probabilities and corresponding return periods varied only slightly over the range of k (as noted above, the number of hours also has little effect on the result). Thus, only single values of the probabilities are given for each form. It is seen that a standard based on the expected maximum value would lead to violations of the present standard average of once every eight years. It is, therefore, significantly more restrictive than the proposed statistical standard at the same concentraiton level.

A standard based on the expected second highest concentration would violate the present deterministic standard on the average of every other year and is, therefore, significantly less restrictive than the proposed statistical form of the standard.

3.4.2 Comparison on the Basis of Fractional Reduction in Precursor Emissions Required to Meet Standard

Another method by which statistical forms can be compared to deterministic forms of a national ambient air standard is to determine the degree of reduction in precursor emissions which each would require. The difficulty discussed at the start of Section 3.4 is particularly significant here. If the present deterministic standard for photochemical oxidants were to be fully enforced, that is, if control levels were required to an extent that assured no exceedances of the standard level, the control levels required would probably be significantly higher than those on which the AQCR's have based their present hydrocarbon control plans. The reason for this is that the design value for deciding level of control required with the present deterministic standard is the second highest concentration. However, the second highest concentration is not characteristic of the general air quality over an AQCR since it is a statistical variable. Even though the levels of all emissions which contributed to oxidant levels over an AQCR were held at a constant level, the annual second highest concentration would vary from one year to the next. While it is true EPA's recommended practice is to have an AQCR select the highest annual second high from several years of data (when available) this does not fully allow for the range of variation which can be experienced in the second high. Thus, any annual second high design value will, to a high probability, lead to undervaluing the degree of hydrocarbon precursor control required.

This problem is resolved with statistical standards. For a statistical standard based on an expected exceedance rate E_j, the design value would be that concentration whose expected exceedance rate is E. This concentration has a fixed value for a given level of contributing precursor emissions and thus is a true characteristic of the general air quality over an AQCR. It can be estimated from a plot of one or more years of ambient air data for a monitoring site. The more years the better the estimate. This estimate will be subject to statistical fluctuation for a finite set of years but will be significantly more stable than the annual second highest concentration.

Given the above mentioned difficulties with the deterministic standard the following comparison was performed: Use was made of data complied by the Monitoring and Data Analysis Division of the Office of Air Quality Planning and Standards which showed observed annual second highest ozone concentraiton for sixty AQCR's having the greatest oxidant problem. The second highest ozone concentrations were distributed as shown in Table 3-14. Accordingly, base case second high concentrations of 0.12, 0.15, 0.20, 0.25, 0.30, and 0.47 were choosen for study to cover the observed range. If it is assumed these values represent the most probable values of the second highest concentrations and the Weibull equation represents the distribution of hourly average concentrations, then each second high corresponds to a given Weibull distribution if the value of k and n are specified. Table 3-14 shows the relative location of the most probable value of the second highest concentration for a range of shape factors, k, and n = 8760.

From the Table 3-14 data the concentrations, C_E, corresponding to E= 2 were calculated for each second high level for a range of k values and n = 8760. Given these data it was then possible to calculate the relative C_E corresponding to each value of E from 1 through 10, 15, and 20 for each k for Weibull distributions corresponding to most probable second highs of 0.12, 0.15, 0.20, 0.25, 0.30 and 0.47 ppm.

Table 3-14 Cumulative Distribution of Second Highest Concentrations for 60 AQCR's

Concentration Second High ppm	Approx. Cum. Distribution %
0.12	10
0.15	25
0.20	50
0.30	90
0.47	Highest

Given the C_E and the second high values for each distribution the percent hydrocarbon control required to get to $C_{STD} = 0.08$ with corresponding number of exceedances was calculated using the proportional roll back method.

$$\% \Delta \text{HC} = \frac{C_E - 0.08}{C_E} \times 100 \tag{3-38}$$

Background is assumed to be zero in Eq. (3-38).

Equation (3-38) is also used to calculate the percent control required under the present deterministic standard by substituting the most probable value of the second highest concentration for C_E. As seen from Table 3-14 the most probable second high is very close to $C_E = 2$ over the range of k. It would, therefore, give about the same percent hydrocarbon control as $C_E = 2$. However, the use of this value alone gives a misleading picture of the second high since, as discussed, it is a statistical variate. To obtain an idea of the range of percent hydrocarbon controls that could be required in an area of a given general air quality high and low values of the second highs (the 5% and 95% vlaues) corresponding to the most probable second high were calculated and used to calculate percent control. The distributions for the annual second highest value were calculated using Eq. (3-26) of Section 3.2.3. Figure 3-8 shows the case for k = 1, n = 8760.

Tables 3-15 through 3-20 show the calculated percent hydrocarbon control and corresponding concentrations for the range of k and n values. It is seen in general that as the allowed number of expected exceedances, E, is increased the percent control required decreases. The fall in percent control as E increases is largest at the lowest concentration levels. The percent control

Figure 3-8

Distribution of Second Highest Value for Exponential Distribution of Hourly Average Concentrations

n=8760

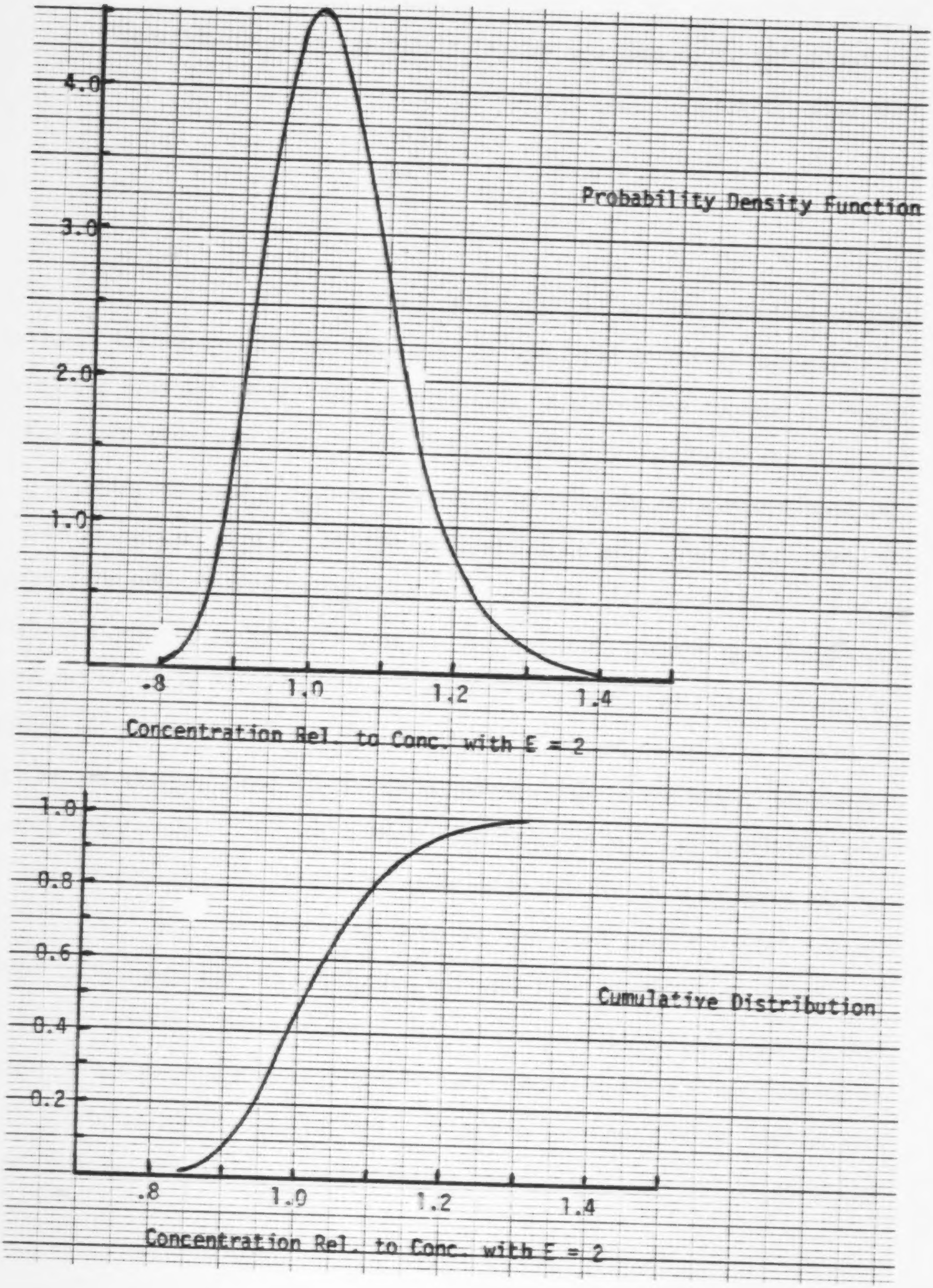


Table 3-15 Location of Most Probable Value of Second Highest Concentration Relative to the Concentration Whose Exceedance Rate is Two (C_2)

Weibull Constant k	Concentration Relative to C_2
0.50	0.9861
0.75	0.9967
1.00	1.0000
1.50	1.0015
2.00	1.0017

n = 8760 hours.

Table 3-16 Change in Percent Control Required to Meet Standard at 0.08 ppm Level with Change in Allowed Expected Exceedances

(Assume Weibull Distribution of Hourly Average Concentrations: $k = 0.5$ $n = 8760$ hrs)

Expected Exceedances per year E	Second Highest Conc.: 0.12 ppm		0.15		0.20		0.25		0.30		0.47	
	C_E(1)	%ΔHC(2)	C_E	%ΔHC	C_E	%ΔHC	C_E	%ΔHC	C_E	%ΔHC	C_E	%ΔHC
1	0.143	44.1	0.178	55.1	0.238	66.4	0.297	73.0	0.356	77.5	0.559	85.7
2	0.122	34.4	0.152	47.4	0.203	60.6	0.253	68.4	0.304	73.7	0.477	83.2
3	0.110	27.6	0.138	41.9	0.184	56.5	0.229	65.1	0.275	70.9	0.432	81.5
4	0.103	22.1	0.128	37.5	0.171	53.2	0.213	62.4	0.256	68.7	0.401	80.1
5	0.097	17.3	0.121	33.7	0.161	50.3	0.201	60.1	0.241	66.8	0.378	78.9
6	0.092	13.2	0.115	30.3	0.153	47.8	0.191	58.1	0.230	65.2	0.360	77.8
7	0.088	9.4	0.110	27.3	0.147	45.5	0.183	56.3	0.220	63.6	0.345	76.8
8	0.085	5.9	0.106	24.5	0.141	43.4	0.176	54.6	0.212	62.2	0.332	75.9
9	0.082	2.6	0.102	21.8	0.137	41.5	0.170	53.0	0.205	60.9	0.321	75.1
10	0.080	-	0.099	19.4	0.133	39.6	0.165	51.6	0.198	59.7	0.311	74.3
15	0.070	-	0.088	8.8	0.117	31.7	0.146	45.2	0.175	54.4	0.275	70.9
20	0.064	-	0.080	-	0.107	25.1	0.133	39.9	0.160	50.0	0.251	68.1
Position of 2nd High Distribution												
Mode	0.120	33.3	0.150	46.7	0.200	60.0	0.250	68.0	0.300	73.3	0.470	83.0
5%	0.0979	18.30	0.1224	34.60	0.1632	50.98	0.2040	60.78	0.2448	67.32	0.3835	79.14
95%	0.1770	54.80	0.2212	63.84	0.2950	72.88	0.3688	78.31	0.4425	81.92	0.6933	88.46

(1) C_E is concentration (ppm) whose expected exceedances per year are E.

(2) %Δ is required % reduction in emissions of nonmethane hydrocarbons by proportional rollback method.

Table 3-17 Change in Percent Control Required to Meet Standard at 0.08 ppm Level with Change in Allowed Expected Exceedances

(Assume Weibull Distribution of Hourly Average Concentrations: $k = 0.75$ $n = 8760$ hrs)

Expected Exceedances per year	Second Highest Conc.: 0.12 ppm		0.15		0.20		0.25		0.30		0.47	
E	C_E(1)	%ΔHC(2)	C_E	%ΔHC	C_E	%ΔHC	C_E	%ΔHC	C_E	%ΔHC	C_E	%ΔHC
1	0.133	40.0	0.167	52.0	0.223	64.2	0.279	71.3	0.335	76.1	0.525	84.8
2	0.120	33.3	0.150	46.7	0.201	60.2	0.251	68.1	0.301	73.4	0.472	83.1
3	0.112	28.8	0.140	43.0	0.188	57.5	0.235	65.9	0.282	71.6	0.442	81.9
4	0.107	25.2	0.134	40.2	0.179	55.3	0.224	64.2	0.268	70.2	0.421	81.0
5	0.103	22.2	0.129	37.8	0.172	53.6	0.215	62.8	0.258	69.0	0.405	80.2
6	0.100	19.6	0.124	35.7	0.167	52.0	0.208	61.6	0.250	67.9	0.391	79.6
7	0.097	17.3	0.121	33.8	0.162	50.6	0.202	60.5	0.243	67.0	0.380	79.0
8	0.094	15.2	0.118	32.1	0.158	49.4	0.197	59.4	0.237	66.2	0.371	78.4
9	0.092	13.2	0.115	30.6	0.154	48.2	0.193	58.5	0.231	65.4	0.363	77.9
10	0.090	11.4	0.113	29.1	0.151	47.1	0.189	57.7	0.227	64.7	0.355	77.5
15	0.083	3.8	0.104	23.1	0.139	42.6	0.174	54.0	0.209	61.7	0.327	75.5
20	0.078	-	0.098	18.2	0.131	38.9	0.164	51.1	0.196	59.2	0.308	74.0
Position of 2nd High Distribution												
Mode	0.120	33.3	0.150	46.7	0.200	60.0	0.250	68.0	0.300	73.3	0.470	83.0
5%	0.1042	23.20	0.1302	38.56	0.1736	53.92	0.2170	63.13	0.2604	69.28	0.4079	80.39
95%	0.1546	48.24	0.1932	58.59	0.2576	68.94	0.3220	75.16	0.3864	79.30	0.6054	86.79

(1) C_E is concentration (ppm) whose expected exceedances per year are E.

(2) %Δ is required % reduction in emissions of nonmethane hydrocarbons by proportional rollback method.

Table 3-18 Change in Percent Control Required to Meet Standard at 0.08 ppm Level with Change in Allowed Expected Exceedances

(Assume Weibull Distribution of Hourly Average Concentrations: k = 1.0 n = 8760 hrs)

Expected Exceedances per year E	Second Highest Conc.: 0.12 ppm		0.15		0.20		0.25		0.30		0.47	
	C_E(1)	%ΔHC(2)	C_E	%ΔHC	C_E	%ΔHC	C_E	%ΔHC	C_E	%ΔHC	C_E	%ΔHC
1	0.130	38.4	0.162	50.7	0.217	63.1	0.271	70.4	0.325	75.4	0.509	84.3
2	0.120	33.3	0.150	46.7	0.200	60.0	0.250	68.0	0.300	73.3	0.470	83.0
3	0.114	29.9	0.143	44.0	0.190	58.0	0.238	66.4	0.285	72.0	0.447	82.1
4	0.110	27.3	0.138	41.9	0.183	56.4	0.229	65.1	0.275	70.9	0.431	81.4
5	0.107	25.2	0.134	40.1	0.178	55.1	0.223	64.1	0.267	70.1	0.419	80.9
6	0.104	23.3	0.103	38.6	0.174	54.0	0.217	63.2	0.261	69.3	0.408	80.4
7	0.102	21.6	0.128	37.3	0.167	52.1	0.213	62.4	0.255	68.6	0.400	80.0
8	0.100	20.1	0.125	36.1	0.164	51.3	0.209	61.7	0.250	68.1	0.392	79.6
9	0.098	18.7	0.123	35.0	0.162	50.5	0.205	61.0	0.246	67.5	0.386	79.3
10	0.097	17.5	0.121	34.0	0.159	49.8	0.202	60.4	0.242	67.0	0.380	78.9
15	0.091	12.2	0.114	29.8	0.152	47.3	0.190	57.9	0.228	64.9	0.357	77.6
20	0.087	8.1	0.109	26.5	0.145	44.9	0.181	55.9	0.218	63.2	0.341	76.5
Position of 2nd High Distribution												
Mode	0.120	33.3	0.150	46.7	0.200	60.0	0.250	68.0	0.300	73.3	0.470	83.0
5%	0.1076	25.68	0.1346	40.54	0.1794	55.41	0.2243	64.33	0.2691	70.27	0.4216	81.02
95%	0.1447	44.72	0.1809	55.78	0.2412	66.83	0.3015	73.47	0.3618	77.89	0.5668	85.89

(1) C_E is concentration (ppm) whose expected exceedances per year are E.

(2) %Δ is required % reduction in emissions of nonmethane hydrocarbons by proportional rollback method.

Table 3-19 Change in Percent Control Required to Meet Standard at 0.08 ppm Level with Change in Allowed Expected Exceedances

(Assume Weibull Distribution of Hourly Average Concentrations: k = 1.5 n = 8760 hrs)

Expected Exceedances per year E	Second Highest Conc.: 0.12 ppm		0.15		0.20		0.25		0.30		0.47	
	C_E (1)	%ΔHC (2)	C_E	%ΔHC	C_E	%ΔHC	C_E	%ΔHC	C_E	%ΔHC	C_E	%ΔHC
1	0.127	36.8	0.158	49.4	0.211	62.1	0.264	69.7	0.316	74.7	0.495	83.8
2	0.120	33.3	0.150	46.7	0.200	60.0	0.250	68.0	0.300	73.3	0.469	82.9
3	0.116	31.1	0.145	44.9	0.193	58.7	0.242	66.9	0.290	72.4	0.454	82.4
4	0.113	29.4	0.142	43.5	0.189	57.6	0.236	66.1	0.283	71.8	0.443	81.9
5	0.111	28.0	0.139	42.4	0.185	56.8	0.231	65.4	0.278	71.2	0.434	81.6
6	0.109	26.8	0.137	41.4	0.182	56.1	0.228	64.9	0.273	70.7	0.427	81.3
7	0.108	25.7	0.135	40.6	0.180	55.4	0.224	64.4	0.269	70.3	0.421	81.0
8	0.106	24.8	0.133	39.8	0.177	54.9	0.222	63.9	0.266	69.9	0.416	80.8
9	0.105	23.9	0.131	39.2	0.175	54.4	0.219	63.5	0.263	69.6	0.411	80.5
10	0.104	23.2	0.130	38.5	0.174	53.9	0.217	63.1	0.260	69.3	0.407	80.3
15	0.100	19.9	0.125	35.9	0.167	52.8	0.208	61.6	0.250	68.0	0.390	79.5
20	0.099	19.4	0.121	33.9	0.161	50.5	0.202	60.4	0.242	67.0	0.379	78.9
Position of 2nd High Distribution												
Mode	0.120	33.3	0.150	46.7	0.200	60.0	0.250	68.0	0.300	73.3	0.470	83.0
5%	0.1114	28.22	0.1393	42.57	0.1857	56.93	0.2322	65.54	0.2786	71.29	0.4365	81.67
95%	0.1358	41.07	0.1697	52.86	0.2263	64.64	0.2828	71.71	0.3394	76.43	0.5317	84.95

(1) C_E is concentration (ppm) whose expected exceedances per year are E.

(2) %Δ is required % reduction in emissions of nonmethane hydrocarbons by proportional rollback method.

Table 3-20 Change in Percent Control Required to Meet Standard at 0.08 ppm Level with Change in Allowed Expected Exceedances

(Assume Weibull Distribution of Hourly Average Concentrations: k = 2.0 n =8760 hrs)

Expected Exceedances per year E	Second Highest Conc.: 0.12 ppm		0.15		0.20		0.25		0.30		0.47	
	C_E (1)	%ΔHC (2)	C_E	%ΔHC	C_E	%ΔHC	C_E	%ΔHC	C_E	%ΔHC	C_E	%ΔHC
1	0.125	35.9	0.156	48.7	0.208	61.6	0.260	69.2	0.311	74.3	0.488	83.6
2	0.120	33.3	0.150	46.7	0.200	60.0	0.250	68.0	0.299	73.2	0.469	82.9
3	0.117	31.7	0.146	45.3	0.195	59.0	0.244	67.2	0.292	72.6	0.458	82.5
4	0.115	30.4	0.144	44.3	0.192	58.2	0.239	66.6	0.286	72.1	0.449	82.2
5	0.113	29.4	0.142	43.5	0.189	57.6	0.236	66.1	0.282	71.7	0.443	81.9
6	0.112	28.5	0.140	42.8	0.186	57.1	0.233	65.7	0.279	71.3	0.437	81.7
7	0.111	27.7	0.138	42.2	0.184	56.6	0.231	65.3	0.276	71.0	0.433	81.5
8	0.110	27.0	0.137	41.6	0.183	56.2	0.228	65.0	0.273	70.7	0.428	81.3
9	0.109	26.4	0.136	41.1	0.181	55.8	0.226	64.7	0.271	70.5	0.425	81.2
10	0.108	25.8	0.135	40.7	0.180	55.5	0.225	64.4	0.269	70.2	0.422	81.0
15	0.105	23.5	0.131	38.8	0.174	54.1	0.218	63.3	0.261	69.3	0.409	80.4
20	0.102	21.7	0.128	37.4	0.170	53.0	0.213	62.4	0.255	68.6	0.399	80.0
Position of 2nd High Distribution												
Mode	0.120	33.3	0.150	46.7	0.200	60.0	0.250	68.0	0.300	73.3	0.470	83.0
5%	0.1137	29.61	0.1418	43.59	0.1891	57.69	0.2364	66.16	0.2837	71.80	0.4444	82.00
95%	0.1318	39.29	0.1645	51.35	0.2193	63.52	0.2741	70.81	0.3289	75.68	0.5153	84.47

(1) C_E is concentration (ppm) whose expected exceedances per year are E.

(2) %Δ is required % reduction in emissions of nonmethane hydrocarbons by proportional rollback method.

required when the most probable value of the second high is used is close to that for $E = 2$. The control required by the 5 and 95 percentile second highs cover a wide range corresponding approximately to $E = 5$ at the 5% value to control levels significantly above that for $E = 1$ at the 95% value. Increasing k causes significant reduction in the effects of increasing E values.

In view of the discussion in Section 3.3 it would have been more correct in the above calculations to use $n = 1071$ for calculating the distributions of the annual second high concentrations. In doing this, however, it would have been necessary to adjust the shape factors, k, to somewhat higher values. Since, as discussed in Section 3.3, this tends to undo the effect of going to a lower k, the calculations were kept simple by using $n = 8760$ and not changing the k values.

3.4.3 The Highest Second Highest Concentration

EPA's recommended practice with respect to the present deterministic photochemical oxidants standard is that where the second high concentration for several years are available, the highest of these values be used. It is, therefore, of interest to determine the most probable value of the annual second highest concentration if more than one year of data are available. Define:

$$F_{(2)} = P(C_{(2)} \leq C), \tag{3-39}$$

where $C_{(2)}$ is the second highest observation of a series of n observations.

If there are N such second highest values, by the binomial distribution it is known that the cumulative distribution of the highest of N second high values is(assuming independence of concentrations):

$$G_{(2)} = F_{(2)}{}^{N} \tag{3-40}$$

The corresponding density function is given by:

$$g_{(2)} = \frac{d\,G_{(2)}}{dC} = NF_{(2)}{}^{N-1}\,\frac{d\,F_{(2)}}{dC} \tag{3-41}$$

From Eq. (3-26) the cumulative distribution of the second highest value in n independent trials is:

$$F_{(2)} = n(1-p)^{n-1}p + (1-p)^{n} \tag{3-42}$$

and

$$\frac{d\,f_{(2)}}{dC} = n(n-1)\,(1-p)^{n-2}\,p(-\,\frac{dp}{dC}) \tag{3-43}$$

Following the practice in Section 3.2.1:

$$p = e^{-(\ln\frac{n}{E})\,(\frac{C}{C_E})^{k}} \tag{3-44}$$

Then:

$$\frac{dp}{cC} = - \frac{k}{C_E} (\ln n?E) \, (C/C_E)^{k-1} \, p \tag{3-45}$$

Therefore,

$$\frac{d\,F_{(2)}}{dC} = n(n-1) \, (\frac{k}{C_E}) \, \ln n/E \, (\frac{C}{C_E})^{k-1} \, (1-p)^{n-2} \, p^2 \tag{3-46}$$

Substitution of Eqs. (3-42) and (3-46) into (3-40) and (3-41) allows the calculation of the cumulative distribution and density function.

Table 3-21 shows the calculated position of the most probable highest second high concentrations for N years relative to $C_{E = 2}$ for a range of k values and n = 8760. As N increases the position of the most probable value shifts to higher concentrations. Table 3-21 is useful in connection with Tables 3-16 through 3-20. It shows, for example, that for k = 1, and $C_2 = 0.12$, the most probable highest second high for N = 10 is 0.12 x 1.1653 = 0.140 ppm. Reference to Table 3-18 shows the 95 percentile second high is 0.1447 ppm which is, as expected, just somewhat higher than the most probable highest second high over 10 years.

The density functions for the highest second highs for k = 1 are shown in Table 3-22.

Table 3-21 Position of Most Probable Highest Second High Concentration of N Years Relative to Concentration with Two Expected Exceedances per Year

N (Years)		Relative Position of Highest Second High Concentration				
	K =	0.5	0.75	1.0	1.5	2.0
		C = 1,	E = 2,	n = 8760		
1		0.9861	0.9967	1.0000	1.0015	1.0017
2		1.0968	1.0688	1.0531	1.0363	1.0276
3		1.1610	1.1094	1.0826	1.0554	1.0417
4		1.2059	1.1373	1.1030	1.0685	1.0514
5		1.2407	1.1587	1.1185	1.0785	1.0587
10		1.3479	1.2239	1.1653	1.1081	1.0804

Table 3-22 Density Function for Distribution of Highest Second Highest Value in N Trials of 8760 Trials Each for Weibull Distribution of Hourly Averages Where k = 1, C_E = 1.0, E = 2

Conc. (ppm)	N = 1	2	3	4	5	10
0.6	0.0000	-	-	-	-	-
0.7	0.0000	-	-	-	-	-
0.8	0.0216	0.0000	0.0000	0.0000	0.0000	-
0.9	1.7572	0.1935	0.0160	0.0012	0.0001	0.0000
1.0	4.5396	3.6859	2.2446	1.2150	0.6166	0.0136
1.1	2.6407	4.1479	4.8864	5.1167	5.0231	3.0016
1.2	0.8064	1.5247	2.1619	2.7250	3.2199	4.8613
1.3	0.1864	0.3684	0.5461	0.7196	0.8890	1.6760
1.4	0.0382	0.0762	0.1140	0.1517	0.1892	0.3740
1.5	0.0074	0.0149	0.0223	0.0297	0.0371	0.0740
1.6	0.0014	0.0028	0.0042	0.0056	0.0071	0.0141
1.7	0.0003	0.0005	0.0008	0.0011	0.0013	0.0027
1.8	0.0000	0.0001	0.0001	0.0002	0.0002	0.0005
1.9			0.0000	0.0000	0.0000	0.0001
2.0						0.0000
Max C	1.0000	1.0531	1.0826	1.1030	1.1185	1.1653
p Max	4.5396	4.8430	5.0097	5.1211	5.2030	5.4299

3.5 Statistical Standard Based on the Daily Maximum Hourly Average Concentration

The discussion in Section 3 has dealt primarily with short-term standards in which the averaging time was one hour. It was pointed out that much of the discussion was also applicable to longer averaging times. The standards based on averaging times all have the common feature of dividing the year into a series of contiguous segments of equal duration. A potentially useful alternate approach is to be only concerned with that period during each day in which the time-averaged concentration reaches its maximum value for the day. The averaging time should be small compared to twenty four hours. For ozone the time would be one hour. A statistical daily maximum hour standard would be, for example:

> 0.08 ppm daily maximum hourly average concentration with expected exceedances per year less than or equal to one.

Or the standard might be based on the maximum permitted value of the expected highest daily maximum hourly average concentration in one year. The standard also could be expressed in a deterministic form.

A daily maximum hour standard would simplify the reporting and analysis of ambient air data. However, it must also be acceptable from the point of view of protecting public health. Since a single exceedance of a daily maximum hour standard in a year may represent one, two, or more exceedances of an hourly standard it is clear a daily maximum standard is less stringent than an hourly standard when both standards are set at the same level. The following sections explore the magnitude of this difference for both statistical and deterministic forms of the standard.

3.5.1 Comparison of Hourly and Daily Maximum Hour Standards When Hourly Average Concentrations Exhibit No Dependence

Comparison of standards based on hourly average concentrations and daily maximum average concentrations is complicated because the lack of independence of one hourly average concentration from another and the generally nonstationary character of the stochastic processes generating ambient concentrations of pollutants.

As a starting point it is worthwhile, however, to see how the two types of standards would compare if complete independence and a stationary stochastic process were assumed. If it is also assumed the distribution of the hourly averages is given by a Weibull distribution, it is possible to derive the relationships between the two standards.

The basic hourly average concentration distribution is:

$$P_h = e^{-(\ln \frac{n_{h/y}}{E_{h/y}})\,(C/C_{HSTD})^{k}} \qquad (3\text{-}47)$$

where:

P_h = $P\,(C_{obs.} \geq C)$

C_{HSTD} = The concentration level of the hourly average standard.

$E_{h/y}$ = The expected number of hours per year exceeding the standard.

$n_{h/y}$ = The effective hours per year.

If each hourly average concentration is independent of the others and the stochastic process is stationary, the probability that the concentration C will not be exceeded in $n_{h/d}$ hours (the effective number of hours in a day) is given by the binomial distribution:

$$p_o = (1-p_h(C))^{n_{h/d}} \qquad (3\text{-}48)$$

This probability is also the probability that the maximum observed hourly average is less than or equal to the concentration C. Therefore:

$$p_d = 1 - (1 - p_h(C))^{n_{h/d}} \qquad (3\text{-}49)$$

is the probability distribution for the daily maximum hourly average concentration.

If a standard concentration level and expected number of exceedances is set for the daily hourly average concentration, can write:

$$\frac{E_{d/y}}{n_{d/y}} = 1 - (1 - p_h\ (C_{DSTD}))^{n_{h/d}} \qquad (3\text{-}50)$$

where:

C_{DSTD} = The concentration level of the daily maximum hourly average standard.

$E_{d/y}$ = The expected number of days per year exceeding the standard.

$n_{d/y}$ = The effective number of days per year.

If p_h is given by (3-47) it follows that:

$$p_h\,(C_{DSTD}) = e^{-\left(\ln \frac{n_{h/y}}{E_{h/y}}\right)\left(\frac{C_{DSTD}}{C_{HSTD}}\right)^{k}} \qquad (3\text{-}51)$$

Substituting Eq. (3-51) into (3-50) yields:

$$\frac{C_{DSTD}}{C_{HSTD}} = \left[\frac{\ln \left[1 - (1- \frac{E_{d/y}}{n_{d/y}})^{1/n_{h/d}}\right]^{-1}}{\ln \frac{n_{h/y}}{E_{h/y}}} \right]^{1/k} \qquad (3\text{-}52)$$

Equation (3-52) connects the parameters of the two standards when both are connected by the same distribution of hourly average concentrations. It allows comparisons to be made between the two kinds of standards when both standards are simultaneously met. Table 3-23 shows the corresponding levels of the standards when the expected exceedances for each are the same. In the one case the hourly average standard level is held constant at 0.08 ppm for each value of the exceedance. In the other, the daily maximum hourly average concentration level is held constant. In this table the effective hours per year are the full 8760 and the effective hours per day are 24. The calculation is done for three levels of k.

It is seen that in the range of 1 to 20 exceedances, the difference between the two standards is negligible for complete independence of hours and a stationary stochastic process.

To partially remove the effect of nonstationarity, the calculation was repeated for effective hours per year of 600 and effective hours per day of 5 (Table 3-24). There is still not much effect up to 10 exceedances.

Thus, for the case of independent hours the two standards are essentially identical for expected exceedances in the range of interest. That also means the risks of exceeding thresholds would be the same. That is, either one could be used for the same concentration level and expected number of exceedances within reasonable limits and give the same protection.

The analysis says that a relatively large number of exceedances would be required before there was sufficient clustering within a day (in this case by a purely random process) to cause the standards to begin to separate.

Unfortunately, the analysis cannot be stopped here because while we have accounted for nonstationarity to some extent we probably have not in this case, gone far enough. For example, suppose a simple, but not far fetched, coupling were introduced, such as: If one exceedance occurs the next hour will almost always exceed the standard, while the probability that the third hour will exceed the standard is about the same as the first hour. In this case the exceedances will cluster very strongly into pairs. The above analysis of the independence case shows that beyond that little additional random clustering will occur if the total exceedances are not too large (< 5).

Now in this case it can be concluded that if the same concentration level is set for the two types of standard the following very nearly holds:

$$E_{d/y} = 1/2\ E_{h/y} \tag{3-53}$$

In other words, the parameters of the two standards would now have to be significantly different to provide the same protection.

The analysis in the preceeding paragraphs shows that for complete independence the standards are nearly the same over a broad range of E values. For ozone concentrations, however, the nonstationarity is extensive. To go further requires working directly with ambient air data since at the present time there is no mathematical model for the nonstationary stochastic processes generating hourly average ozone concentrations.

Table 3-23 COMPARISON OF DAILY MAXIMUM HOURLY AVERAGE AND HOURLY AVERAGE STANDARDS(1)

Expected Exceedances(2)	Level of Daily Maximum Standard When Hourly Standard Held at 0.08 ppm		
k =	0.5	1.0	2.0
1	0.0800	0.0800	0.0800
5	0.0799	0.0799	0.0800
10	0.0797	0.0798	0.0799
15	0.0795	0.0797	0.0799
20	0.0793	0.0796	0.0798
50	0.0779	0.0789	0.0795

	Level of Hourly Standard When Daily Maximum Standard Held at 0.08 ppm		
k =	0.5	1.0	2.0
1	0.0800	0.0800	0.0800
5	0.0801	0.0801	0.0800
10	0.0803	0.0802	0.0801
15	0.0805	0.0803	0.0801
20	0.0807	0.0804	0.0802
50	0.0822	0.0811	0.0805

(1) Number of hours per year = 8760

(2) Number of hours per day = 24

Expected exceedances the same for both standards.

Table 3-24 COMPARISON OF DAILY MAXIMUM AVERAGE AND HOURLY AVERAGE STANDARDS WITH PARTIAL CORRECTION FOR NONSTATIONARITY IN HOURLY AVERAGE CONCENTRATIONS(1)

Expected Exceedances(2)	Level of Daily Maximum Standard When Hourly Standard Held at 0.08 ppm		
k =	0.5	1.0	2.0
1	0.0799	0.0800	0.0800
5	0.0794	0.0797	0.0799
10	0.0787	0.0793	0.0797
15	0.0777	0.0789	0.0794
20	0.0767	0.0783	0.0792
50	0.0674	0.0734	0.0766

	Level of Hourly Standard When Daily Maximum Standard Held at 0.08 ppm		
k =	0.5	1.0	2.0
1	0.0801	0.0800	0.0800
5	0.0806	0.0803	0.0801
10	0.0814	0.0807	0.0803
15	0.0823	0.0812	0.0806
20	0.0835	0.0817	0.0809
50	0.0950	0.0872	0.0835

(1) Effective number of hours per year = 600
Effective number of hours per day = 5

(2) Expected exceedances the same for both standards.

3.5.2 Comparison of Hourly and Daily Maximum Standards When Hourly Average Concentrations Exhibit Dependence

To determine the effect of dependence of hourly average concentrations use was made of an analysis performed by the Monitoring and Data Analysis Division of OAQPS, at the Pollutants Strategies Branch request on data from 93 urbanized areas with populations greater than 200,000. Data were assembled on the three highest ozone readings for selected sites in each area and the dates on which they occurred. In a number of cases the highest readings would duplicate and thus lead in some cases to having, in effect, as many as the six highest readings. The dates of each of these readings were also recorded. From these data it could be determined whether or not the second highest day and the second highest hour tend to correspond to the same reading. The MDAD data are shown in Table 3-25 grouped according to air quality region. The right hand column shows that in 55% of the site-years the second highest hour and the second highest daily maximum hour are the same. That is, in 55% of the cases the highest hourly average concentration and the second highest concentration occur on different days. In 45% of the cases the 3rd, or 4th, or 5th, etc. highest hourly average corresponds to the second high day.

It is seen in Table 3-25 that the data for each region are shown in two groupings. This is necessary because in many cases insufficient data were available to determine which hourly average concentration corresponded to the second high daily maximum hourly average. For example, if the three highest concentrations were all different, thru data points would have been recorded for that site that year. If these all occurred on the same day then all that would be known is that the second high day corresponded to a

Table 3-25 Distribution of Annual Six Highest Hourly Average Concentration By Region

Region	1	2	3	4	5	6	7	8	9	10	All
% Site-years With Second High Day =											
2nd high hour	52.0	43.2	46.2	55.6	51.0	49.3	50.6	75.7	63.3	65.6	54.65
3rd " "	27.4	20.4	24.9	19.2	24.7	21.1	19.3	16.2	24.1	12.5	22.80
4th " "	2.9	4.6	7.6	5.3	3.6	2.8	3.6	2.7	4.2	6.2	4.56
5th " "	0	2.3	2.7	1.3	0.5	1.4	1.2	0	1.7	0	1.35
6th " "	0	0	0	0	0	0	0	0	0	0	0
Exclusive of Above Site-years % Site-years Where Second High Day Less Than											
3rd hour printed	11.8	25.0	12.5	13.9	13.9	21.1	14.5	5.4	4.6	0	11.49
4th " "	2.9	4.6	1.6	2.0	4.6	4.2	6.0	0	1.0	12.5	2.96
5th " "	2.9	0	1.6	2.6	1.0	0	1.2	0	0.3	0	1.18
6th " "	0	0	2.7	0	0.5	0	3.6	0	0.7	3.13	1.01
Site-years	102	44	184	151	194	71	83	37	286	32	1184

concentration less than the third highest hour. It is possible to redistribute these cases into the upper group if it is assumed the distribution of the upper group is approximately correct. For example, looking at the far right column of Table 3-25, 11.49% indefinite cases have the second high day corresponding to less than the third highest hour. These cases can be distributed among the 4th, 5th, 6th highest, etc. in the grouping above roughly in proportion to the current distribution.

Thus, if the 11.69 is assumed to distribute entirely between the 4th, and 5th high, 11.49 x 4.56/5.91 = 8.9% should go to the 4th and 11.49 x 1.35/5.91 = 2.6% should go to the 5th high. The 2.96% less than the fourth high (Table 3-25, far right column) can be distributed to the 5th high. The 1.2% greater than the 5th high can be distributed to the sixth high. The 1% remaining is attributed to the seventh and higher values. The calculation could be repeated using the newly obtained distribution. Table 3-26 shows the result for a single iteration. Using these data to weight the ranking it can be calculated (Table 3-26) that the second highest daily maximum concentration corresponds on the average to the 2.8 highest hourly concentration in a given year.

Table 3-26 can be used to obtain an approximate comparison of at least one statistical form of the hourly and daily maximum hour forms. However, it is first noted that if the forms of the standards were the expected values of the annual maximum hourly average concentration and daily maximum hourly average concentration, both standards would be identical. This follows from the fact that the maximum hour in any year is also the maximum daily maximum hour. This would not be true of the expected value of the second highest

Table 3-26 Daily Maximum Hourly Average Standard versus Hourly Average Standard

Second Highest Daily Maximum Concentration the Same as:	Per Cent of Years
Second highest hourly concentration	55
Third highest hourly concentration	23
Fourth highest hourly concentration	13
Fifth highest hourly concentration	7
Sixth highest hourly concentration	1
≤ Seventh highest hourly concentration	1
	100

Weighted Average Rank =

$$0.55x2 + 0.23x3 + 0.13x4 + 0.07x5 + 0.01x6 + 0.01x7 = 2.8$$

hour versus the expected value of the second highest daily maximum hour. The equivalent daily maximum standard would have a lower expected second high concentration.

Where Table 3-26 can be used is in the case of a standard of the form, X ppm with an expected number of exceedances per year less than or equal to one. At any given number of hourly exceedances of a given concentration level the data in Table 3-26 can be used to estimate the distribution of exceedances of the same concentration level by the daily maximum hour concentrations. When the annual hourly exceedances are 0 or 1, the connection is straightforward. If there are no hourly exceedances, there can be no exceedance of a daily maximum hour standard with the same specified concentration level. If there is a single hourly exceedance there will be a single daily maximum hour exceedance.

In the case of two hourly exceedances, the data in Table 3-26 indicate that in 45% of the years there will be a single exceedance of a daily maximum standard. In the remaining 55% of the years there will be two exceedances. Unfortunately, the data in Table 3-26 are not sufficient to fill out the table for more than two exceedances of the hourly average standard. The values shown in Table 3-27 for 3 and 4 hourly exceedances are coarse estimates. The numbers in the single daily exceedance column, however, do follow from Table 3-26. For example, if there are three hourly exceedances but only one daily maximum hour exceedance, this corresponds in Table 3-26 to cases in which the second highest day corresponds to the 4th, the 5th, etc. highest hour. These add to 22% of the years. Similarily, for four hourly exceedances the number is 9%. These have been rounded to 20% and 10% in Table 3-27.

Table 3-27 Distribution of Exceedances per Year for Daily Maximum Hour Standard Versus Exceedances per Year for Hourly Standard(1)

Exceedances per Year of Hourly Standard	Estimated Distribution of Exceedances for Daily Maximum Hour Standard					Weighted(2) Average Exceedances Daily Std.	Relative(3) Frequencey of Exceedances Hourly Std.
	Exceedances per year = 0	1	2	3	4		
0	1.00					0	0.37
1	0.00	1.00				1.00	0.37
2	0.00	0.45	0.55			1.55	0.18
3	0.00	0.20	0.35	0.45		2.25	0.06
≥ 4	0.00	0.10	0.30	0.30	0.30	2.71	0.02

(1) Both standards have the same specified concentration level.

(2) For two hourly exceedances the weighted average daily maximum hour exceedance rate is 0.00 x 0 + 0.45 x 1 + 0.55 x 2 = 1.55.

(3) For expected exceedances/yr = 1, assuming independence of hours.

From this distribution the weighted average exceedances of the daily maximum hour standard can be calculated for each hourly exceedance. A sample calculation is shown in Table 3-27.

To compare the two standards, the next step is to assume a probability distribution of the hourly exceedances. The last column of Table 3-27 shows how they are distributed for an expected exceedance rate of one if all hours of the year are independent (calculated from the binomial distribution). The corresponding expected expected exceedance rate for the daily maximum hourly average is obtained by combining the last two columns of Table 3-27:

$$0.37 \times 0 + 0.37 \times 1.00 + 0.18 \times 1.55 + 0.06 \times 2.25 + 0.02 \times 2.71 = 0.84$$

Thus, the daily standard is only somewhat more restrictive than the hourly standard when the hourly expected exceedance rate is one.

Without much loss of accuracy it can be said that a daily maximum hour standard with an expected exceedance of once per year corresponds to an hourly standard at the same specified concentration level with an expected rate of 1/0.84 = 1.2 exceedances per year.

A sensitivity analysis of the assumptions in Table 3-27 with regard to the distribution corresponding to hourly exceedances of 3 and 4 and the relative frequencies of exceedances of the hourly standard shows that the last figure given is most probably in the range 1.2 to 1.4. If it is assumed 1.3 expected exceedances per year is the most likely value, Eq. (3-14) can be used to calculate the equivalent concentration level of the daily maximum hour standard if both standards have an expected exceedance rate of one. The comparison is shown in Table 3-28.

Table 3-28 Approximate Comparison of Statistical Daily Maximum Hour and Hourly Standards When Both Have Expected Exceedances per Year of One[1]

Level of Hourly Standard (ppm)	Level of Daily Max. Hour Standard (ppm)
0.06	0.061
0.08	0.082
0.10	0.102
0.12	0.123
0.14	0.143

---

(1) Assumes distribution of hourly average concentrations is represented by a Weibull distribution for 8760 hours whose shape factor is: $k = 1.25$.

While the above calculation is crude the conclusion that the two forms of the standard do not differ much in their relative severities and that the daily maximum hour standard is the less severe of the two can be made with reasonable confidence.

3.5.3 Comparison of Deterministic Hourly and Daily Maximum Standards

The preceeding sections compared statistical forms of hourly and daily maximum standards. For a standard in which only one exceedance per year of specified concentration level is permitted the results are different and depend upon how strictly the standard is enforced. If the restriction to no more than one exceedance a year is strictly applied then both standards are identical in their end result since both (following arguments presented in Section 3.1) will lead to an air quality in which the standard level is never exceeded.

In getting to this air quality level, the "design values" for the daily maximum hour standard will initially average lower than those for an hourly standard. This follows from the data and analysis presented in Section 3.4 and 3.5.2. In application of a "not to be exceeded more than once" standard the design value for calculating required levels of emission control is the highest second high over a period of several base years. If only one base year is available, the second high for the area for that year is used. In this case the highest second highest daily maximum hourly average concentration will be equal to less than that of the highest second highest hourly average concentration. This situation is seen in Table 3-29 which was obtained from the MDAD analysis referred to in the previous section and which provided the basis for Tables 3-25 through 3-27. In this case the highest

Table 3-29 Comparison of Percent Hydrocarbon Control Required by Daily Maximum Hourly Average and Hourly Average Standards

Highest Hourly Average Concentration (ppm)	% Control Hourly Max. - % Daily Max.		
	Average (%)	Maximum (%)	Minimum (%)
0.115	1.0	2.8	0
0.144	4.0	11.2	0
0.174	7.0	16.0	0
0.203	4.5	13.0	0
0.233	4.0	10.0	0
0.262	2.8	8.0	0
0.292	1.5	6.0	0
0.321	1.4	5.0	0
0.351	1.4	4.0	0
0.380	1.4	3.5	0

second high for the hourly and daily maximum hour concentration for each of the 93 urbanized areas was used to calculate the percent control of hydrocarbon emissions needed to meet the standard. Linear rollback was used (Eq. (3-38)). The results were summarized to give the average difference in calculated control level as well as the highest and lowest individual value over a range of air quality levels (as exemplified by the highest hourly average concentration).

It is seen in Table 3-29 that at each level of air quality the minimum difference is zero. And indeed, for many of the regions the second highest hour and the second highest daily maximum hour are the same. This follows directly from the results in Table 3-25 which show that this situation occurs on the average in 55% of the site years of data. Thus, for any region as data are collected over an increasing number of years the design values for the two standards will eventually come together.

Thus, in the short term a deterministic daily maximum will either require the same level or less control than an hourly standard. In the long run both standards will require the same level of control.

Notice the situation is different in statistical standards and depends on the form. They are the same for a standard which specifies a maximum value for the expected value of the annual highest concentration. However, for an expected second high standard or standard that sets an expected exceedance rate of a specified level the daily maximum hour standard will be less than the hourly standard if the hourly average concentrations are autocorrelated. In the case of ozone there is a relatively small difference between the standards based on expected exceedances.

3.6 Conclusions and Recommendations

The present deterministic form of the photochemical oxidants standard does not take into account the statistical behavior of hourly average ozone concentrations. As a result the only way in which strict compliance with the standard can be achieved is by controlling precursor emissions to the degree that the standard concentration level is never exceeded. The standard is, therefore, significantly more stringent than apparent from its seeming allowance of at least one exceedance of the standard level per year. The uses of the second highest ozone concentration as an indicator of how far a given Air Quality Control Region is out of compliance and as a design value to calculate required degree of control of precursor emissions are also misleading. Because second high concentrations are statistical variables they do not provide a measure of the overall air quality for a region and will lead to an under estimation of the degree of control of precursor emissions required to gain compliance. Use of the highest second high for several years of data only partially corrects for this problem.

Statistical standards provide a way out of these difficulties by fully taking into account the statistical behavior of air pollutant concentrations. They provide a clearly defined target for determining compliance and a meaningful measure of existing air quality.

The statistical forms unlike deterministic forms lend themselves to quantitative or mathematical treatment. The terms used in statistical standards such as expected exceedance rate or expected annual maximum hourly average concentration have precise meanings and are relatable to the cumulative frequency or probability distributions of the time average concentrations. With some statistical forms, such as expected annual maximum concentration, it is necessary to take into account autocorrelation between concentrations close to each other in time and nonstationarity in the underlying stochastic processes giving rise to the random changes in concentrations. A simple approximate method has been presented dealing with this lack of independence.

Comparison of statistical and deterministic standards which are set at the same concentration level indicates the deterministic standards to be more stringent if the deterministic standard is strictly enforced. An AQCR just meeting a statistical standard allowing an expected exceedance rate of one will violate on an average of once every four years a deterministic standard allowing not more than one exceedance per year of the same concentration level. A statistical standard based on an expected annual maximum concentration will lead to violations of the above deterministic standard on an average of once every eight years when the statistical standard is just met.

When the comparison is made in terms of degree of precursor emission control, in the short term the deterministic standard could appear on the average to be somewhat more lenient. In the long term and if the deterministic standard is strictly enforced it will require greater control of precursor emissions than comparable statistical standards. (In the case of the present deterministic oxidant standard the comparable statistical standard is based on an expected exceedance rate of one. A standard based on one expected annual maximum hourly average concentration would be somewhat, though not strictly, comparable.)

Comparisons between oxidant standards based on hourly and daily maximum concentrations have led to the following conclusions. If both standards are deterministic in the short term it will appear the daily maximum standard is somewhat more lenient than the hourly standard. The difference in apparent control of precursor emissions required for individual Air Quality Control Regions (AQCR), in which the second high design value is based on one to several years of monitoring data, will vary from zero to somewhat over 15%. In the long run both standards are equally stringent and will require the same level of control.

In the case of statistical forms there would be no difference in the two forms if the hourly average concentrations were not autocorrelated. The autocorrelation in the case of oxidant causes the daily maximum standard to be slightly less stringent.

It is recommended that EPA consider changing all its short-term national ambient air quality standards from the present deterministic to suitable statistical forms. Its standards based on annual mean concentrations could be considered as already being in a statistical form.

Initial guidlines have been developed by the Office of Air Quality Planning and Standards on how to determine compliance and degree of departure from compliance with the proposed statistical standard for oxidant. These guidlines are a good first step, but more study is needed on these questions to fully utilize the benefits of the statistical form.

It would also be desirable to perform a more complete investigation of the applicability of Weibull and other distribution functions to ozone and other pollutants. The question of within year and year to year (or longer term) variations in air pollutants requires further study.

Another major area needing investigation is the autocorrelation of air pollutant concentrations and nonstationarity in the underlying stochastic processes. Improved understanding in this area will allow consideration of a broader range of statistical forms and more accurate treatment or aerometric data.

3.7 References

1. U.S. Code of Federal Regulations, 40 CFR, Part 51.14, 1975.

2. "Uses, Limitations, and Technical Basis for Procedures for Quantifying Relationships between Photochemical Oxidants and Precursors", U.S. Environmental Protection Agency, Research Triangle Park, N.C., EPA 450/2-77-021a, November, 1977.

3. "Investigation of Alternative Methods of Expressing the National Ambient Air Quality Standard for Photochemical Oxidants", Final Report on Task #4 of U.S. Environmental Protection Agency Contract No. 68-02-2393, William F. Biller, East Brunswick, N.J., in press.

4. R. I. Larsen, "A Mathematical Model for Relating Air Quality Measurements to Air Quality Standards", U.S. Environmental Protection Agency, Research Triangle Park, N.C., Publication No. AP-89, November, 1971.

5. T.C. Curran and N.H. Frank, "Assessing the Validity of the Lognormal Model When Predicting Maximum Air Pollution Concentrations", Presented at 68th Annual Meeting of the Air Pollution Control Association, Boston, June 15-20, 1975.

6. "The Validity of the Weibull Distribution as a Model for the Analysis of Ambient Ozone Data", Report, U.S. Environmental Protection Agency, PEDCo Environmental, Inc., Cincinatti, Ohio, in press.

7. J. Horowitz and S. Barakatz, "Statistical Analysis of the Maximum Concentration of an Air Pollutant: Effects of Autocorrelation and Nonstationarity", submitted for publication to Atmospheric Environment.

8. J. V. Uspensky, Introduction to Mathematical Probability, McGraw-Hill, 1937, p.46.

9. Private Communication, T. Johnson, C. Nelson, PEDCo Environmental, Inc. Durham, N. C.

4.0 ASSESSMENT OF RISK ASSOCIATED WITH POSSIBLE CHOICES OF A PRIMARY NATIONAL AMBIENT AIR STANDARD

4.1 Background

If National Ambient Air Standards are expressed in statistical forms it is now possible as discussed in the preceeding sections to associate standard levels in a meaningful way with the general levels of air quality which will meet the standard. The way is then paved for beginning to assess the risks of adverse health effects to the most susceptible members of the population associated with alternative levels of the standard.

A quantitative assessment of these risks can be of major value to the EPA standard setting process. This section present a mathematical formalism for risk assessment. It is based on underlying concepts developed by Mr. Thomas Feagans of the Office of Air Quality Planning and Standards. This report will deal only with the mathematical treatment which was the contractor's contribution. A complete discussion of the method may be found elsewhere[4].

4.2 Underlying Principles of Risk Assessment Method

A broad perscription for setting primary National Ambient Air Quality Standards is given by the following passage from the Clean Air Act.

> National primary ambient air quality standards... shall be ambient air quality standards, the attainment and maintenance of which in the judgment of the Administrator, based on [air quality] criteria and allowing an adequate margin of safety, are requisite to protect the public health[1].

Support material to the Act clarifies this perscription somewhat:

> An ambient air quality standard therefore, should be the maximum permissible ambient air level of an air pollution agent or class of such agents (related to a perido of time) which will protect the health of any group of the population[2].

A fundamental concept implicit in the above passages is the concept of a health effects threshold; i.e., a concentration C_T such that if the concentration of pollutant, averaged over a suitable time period, is less than C_T, the pollutant will neither cause nor contribute to effects which are detrimental to human health - even in people who are most susceptible to the health effects at higher concentration levels.

The provisions in the Act for an adequate margin of safety implicitly acknowledges that the concentration of the health effects threshold will be uncertain. That we do not have precise knowledge on the location of health effect thresholds is amply demonstrated in the EPA criteria documents which discuss in detail the health effects research and which are intended to help decide the pollutant concentrations at which health effects occur. In the case of photochemical oxidants, for example, numerous toxicological, clinical, and epidermiological studies are cited which show effects at various levels but do not directly allow a determination of the lowest concentration level of oxidants in ambient air at which the most susceptible portion of the population would suffer adverse health effects. The data from different studies are often in conflict and in cases of effects shown in animals it is not certain the effects observed will be caused to occur in man exposed to oxidants in ambient air.

Fortunately, through the use of techniques developed in the field of decision analysis, the criteria document can be used to quantify the uncertainty in the location of health effect thresholds. Roughly the procedure would be to have highly qualified medical people study the criteria document in detail. After they have absorbed all the pertinent health data, the technique of probability encoding would be used to quantify their subjective judgments by eliciting a probability distribution for the location of individual health effects. That is, they would define functions:

$$D_{T_i} = Pr\ (C_{T_i} \leq C) \qquad (1)$$

Where the quantity D_{T_i} is the probability that the health effects threshold of an effect i is less than or equal to the concentration C. The method of probability encoding by which the probability distribution represented by Eq. (4-1) is determined is discussed in detail in a comprehensive review article[3] and its application to photochemical oxidants is discussed in the report referred to above[4]. The following discussion will take Eq. (4-1) as its starting point and not discuss how it is obtained nor how the concept of health effects threshold is given precise meaning so that it may be successfully encoded. Nor will it be concerned with the concept of subjective probability and its utility in decision making processes. These matters are sufficiently treated in the references cited above.

There is also a second source of uncertainty that must be accounted for in setting an adequate margin of safety. This is the uncertainty in the peak levels that will be reached over a time period of one or more years when the standard is met. As discussed in Section 3, setting National Ambient Air Quality Standards cannot place an absolute lid on the highest pollutant concentrations which may be observed in a given time period. This follows from the stochastic behavior of pollutant concentrations, and,as also discussed in Section 3, leads to the conclusion that standards should be expressed in a way that acknowledges this fact. That is, they should be expressed in suitable statistical forms. Since it is not known what the highest concentration values will be in any given period, the margin of safety again must be adequate to give assurance that these highest concentrations will not be likely to exceed a health effects threshold for the pollutant. The uncertainty in the peak concentration levels can be expressed as:

$$P_C = Pr\ (C_{Max} \leq C) \qquad (4\text{-}2)$$

where the quantity P_C is the probability that no concentration will exceed the concentration level C in a given time period (e.g., 5 years). The concentration C_{Max} is the maximum time averaged concentration observed during the given time period.

It should be noted here that the concentration C appearing in Eqs. (4-1) and (4-2) is a time averaged concentration. The averaging time is set by health effect considerations. If short-term peaks cause the onset of health effects, then averaging times of one hour are used. If exposure to a concentration level over a more extended period of time is necessary, the averaging time could be several hours, one or more days or even a year. For photochemical oxidants the averaging time is, as mentioned earlier, one hour.

To assure that there is an adequate margin of safety is the same as setting the primary standard so that the risk is adequtely small that the most susceptible members of the population will be exposed to concentration levels exceeding a health effect threshold in some given time period. If the expressions (4-1) and (4-2) are known either as mathematical expressions or in tabular form this risk can be calculated for each alternative standard setting. The essential judgment to be made by EPA decision makers is, therefore, deciding on the level of risk that will be tolerated, taking into account the seriousness of the health effects involved. Choosing the acceptable risk level will then lead to choosing the corresponding alternative standard, and this standard will, by definition, have an adequate margin of safety.

A decision also needs to be made regarding the time period. This subject is discussed more completely elsewhere[4]. In brief, it appears reasonable that the time period minimally be keyed into the five year standard review cycle.

4.3 Mathematical Description of the Method

In the preceeding section it was pointed out that the risk of exceeding a true health effects threshold in a given period of time is determined by the uncertainty in the location of the health effects threshold (or in the case of multiple thresholds, the uncertainty in the location of the lowest threshold) and the uncertainty in the maximum oxidant concentration over the given time period. It can be shown, by application of the theory of probability, that the following equation gives the relationship between the risks and the above uncertainties:

$$R = 1 - \int_0^\infty P_C(C) p_T(C) dC \qquad (4\text{-}3)$$

where:

R = Probability (Risk) that a true health effect threshold, or the lowest of a multiple number of thresholds, is exceeded one or more times in a given time period (e.g. one year, five years, etc.).

$P_C(C)$ = Probability that the highest observed time averaged (e.g. one hour, two hour, etc.) pollutant concentration does not exceed the concentration C in the given time period.

$P_T(C)$ = The probability density function for the health effect threshold or in the case of multiple health effects the function for the lowest effect (the composite density function).

The derivation of Equation (4-3) is given in Appendix B.

Equation (4-3) also holds for a more general case in which R is defined as the risk of m or more exceedances of a threshold in a given time period, where m may have the integral values 1, 2, 3, etc. In this case P_C is redefined as the probability that the mth highest time averaged pollutant concentration does not exceed the concentration C in the given time period. In other words, P_C is the cumulative distribution of the mth highest time averaged pollutant concentration in the time period of interest. Specifying the national ambient air quality standard for a pollutant limits the range of $P_C(C)$ functions which will satisfy the air quality requirements of the standard. If this limitation can be expressed quantitatively, and the health effect threshold density functions have been determined, then for any given specification of the standard, a range of risks associated with that specification can be calculated from Eq. (4-3).

In practice, it is only convenient to use Eq. (4-3) directly to calculate risk when a single health effect with an existence probability of one is involved. When the risk that lowest of n health effects will be exceeded is to be calculated and when there is uncertainty as to whether one or more of the effects actually occur in the sensitive population at any attainable pollutant concentration, the following expanded version of Eq. (4-3) is used. (See Appendix B):

$$R = 1 - \int_0^\infty P_C(C) p_T^\circ \, dC - (1 - e_1)(1 - e_2) \cdots (1 - e_n) \qquad (4\text{-}4)$$

where:

p_T° = The probability density function for the location of the lowest of n thresholds over the possible range of concentrations of the pollutant.

e_i = The probability that the ith health effect actually occurs in the most sensitive population in the possible range of concentrations of the pollutant.

The function p_T° is calculated from:

$$p_T^\circ = Q \left[\frac{e_1 p_1^\circ}{Q_1} + \frac{e_2 p_2^\circ}{Q_2} + \cdots + \frac{e_n p_n^\circ}{Q_n}\right] \qquad (4\text{-}5)$$

where:

p_i° = The probability density function for the threshold of the ith health effect assuming its $e_i = 1$. (That is,

$$\int_0^\infty p_i^\circ \, dC = 1).$$

In practice p_i° is obtained by asking the subject health expert in an encoding session to first give his best judgment of the value of e_i and then encoding him as to the location of the threshold assuming that the effect actually occurs in the sensitive population at an attainable pollutant concentration. The encoding procedure for the location of health effects threshold gives the cumulative distributions for the health effects thresholds. These functions are differentiated by numerical methods to obtain the p_i°.

The terms Q_i and Q in Eq. (4-5) are calculated from:

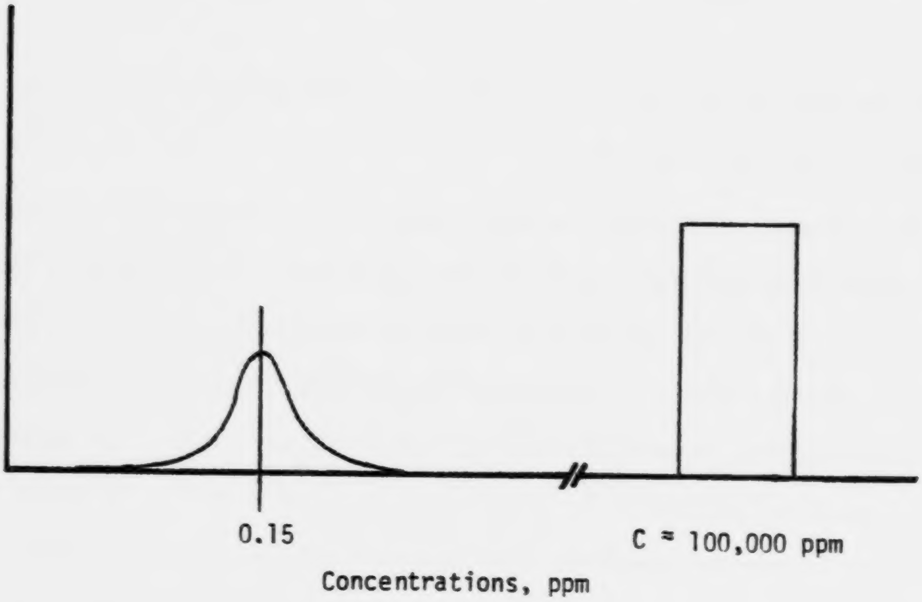
$$Q_i = 1 - e_i \int_{0}^{C} p_i^{\circ} \, dC \qquad (4\text{-}6)$$

$$Q = Q_1 Q_2 Q_3 \cdots Q_n \qquad (4\text{-}7)$$

The Q_i are seen to be functions of the cumulative distribution of the probability density functions p_i°.

The approach used above in treating uncertainty about the existence of a causal relationship between the pollutant and a given health effect is to define a density function p_T° assuming the effect exists and a probability e that the effect is caused by the pollutant to occur in humans. The area under the probability density functions p_T° is one. The approach multiplies this function by e, thus reducing the area to e. To bring the total area to one the remaining area (1-e) is assumed to be accounted for in a concentration range very much above the concentration range which effectively contains p_T°. This situation is illustrated in Fig. 4-1. The distribution on the left is ep°. The area under this function is, therefore, e. The function to the right can have any shape. However, the area under the function must be (1-e). A more detailed treatment of the approach is given in Appendix B.

Figure 4-1 Hypothetical Threshold Probability Density Function for a Health Effect with a 20% Chance of Being Real

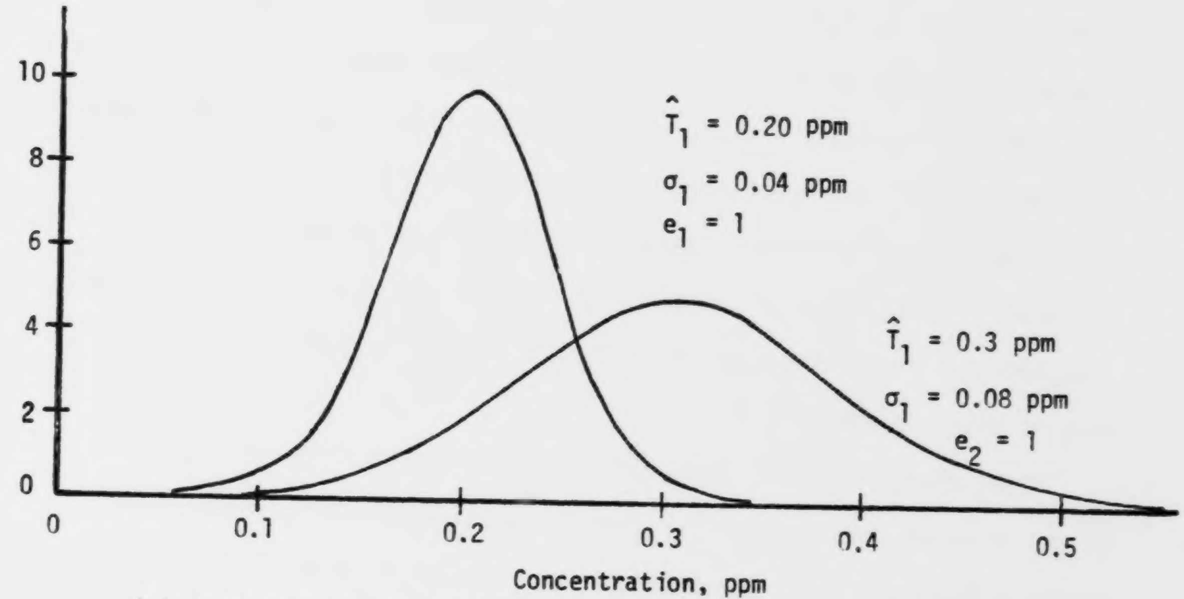


When more than one independent health effect is involved, p_T° is the distribution of the threshold for the lowest effect. This distribution is a composite of the distributions of the individual thresholds and is formed from them in accordance with Eq. (4-5). Figure 4-2(a) shows two threshold distributions for two hypothetical, independent health effects. Figure 4-2(b) shows three different composite distributions formed by assigning different probabilities of existence to the two health effects. A further discussion of compositing and the effect of assigning existence probabilities is given in Appendix C.

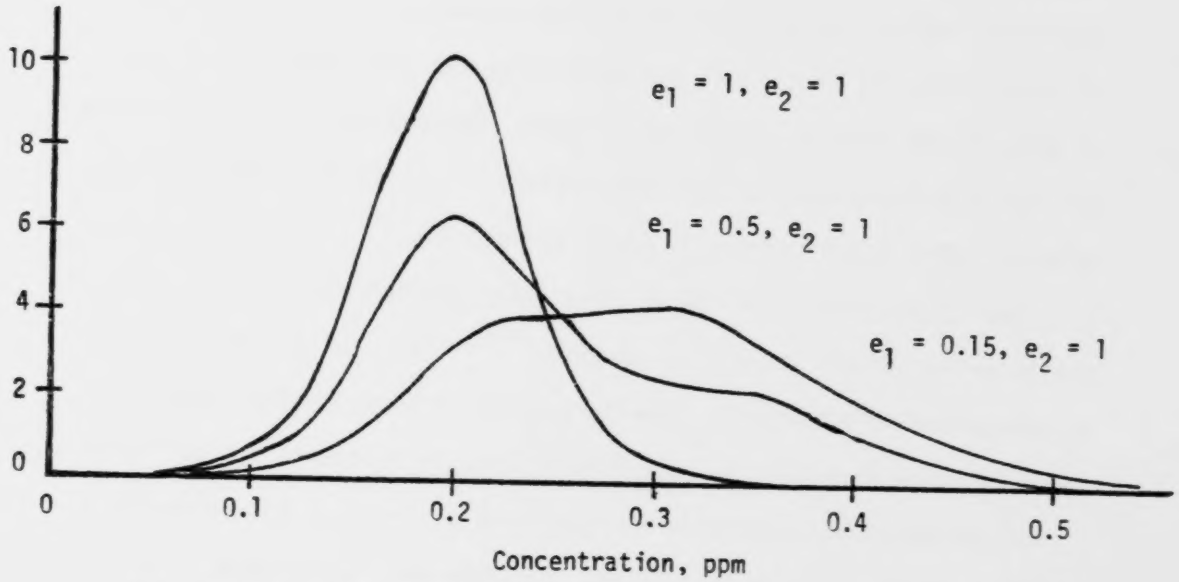
Thus, given the $P_C(C)$ functions corresponding to different levels of the standard, the probability density functions p_i° for n independent health effects and their corresponding e_i, Eqs. (4-4) through (4-7) can be used to calculate the range of risks associated with alternate specifications of the ambient air quality standard. The risks can be calculated for the individual health effects and for composites of two or more of the effects in any combination. Calculating risks for individual health effects and various combinations can be of value where the effects differ significantly in their seriousness.

The calculations involved in Eqs. (4-4) through (4-7) can be most conveniently carried out with a computer. The function p_T° can also be calcualted by differentiating its cumulative distribution function. The cumulative distribution is a function of the existence probabilities e_i and the cumulative distribution functions of the p_i°. (See Appendix B.) Less computational labor is involved with this method if there is no interest in knowing the density functions p_i°.

Figure 4-2 Variation in Composite Health Effect Threshold Probability Density Functions of Two Independent Density Functions as Different Probabilities of Existence are Assigned to the Lower Threshold Function



(a) Individual Health Effect Threshold Probability Density Functions



(b) Composite Probability Density Functions for Various Probabilities of Existence are Assigned to Function with $\hat{T}$ = 0.2 ppm

4.4 Obtaining the $P_C(C)$ Distributions

As indicated in earlier sections the $P_C(C)$ function is the cumulative distribution function for highest time averaged pollutant concentration for a specified period of time. If the risk is calculated for m or more exceedances, it is the cumulative distribution for the mth highest concentration. To simplify the following discussion it will be assumed that the concentration averaging time is one hour.

The $P_C(C)$ function for a pollutant in the air over a given region is a measure of the air quality for that region with respect to the pollutant. Specifying a National Ambient Air Quality Standard places a limitation on the range of $P_C(C)$ functions corresponding to air quality just meeting the standard. For example, if the ambient air quality standard specified an expected (average) value of maximum hourly average concentration for a one year period was not to exceed a given level, this would immediately locate the mean value of the distribution of maximum values and thus define the concentration region in which the preponderance of maximum values must occur. However, depending upon the area and the control methods used to meet the standard, the distribution of maximum values about the mean could be relatively narrow or spread out. It is expected, however, that there would be practical limitations on the degree of spread of the distribution. Therefore, specifying the expected maximum concentration limits the $P_C(C)$ distributions just satisfying the standard.

In applying the risk assessment method it is necessary to determine the range of $P_C(C)$ functions just meeting each alternate specification of the standard. While, in principle, this should be possible for almost any type of standard it is more readily done for standards with statistical forms than for standards with deterministic forms.

The P_C distribution function is related to the distribution of hourly average concentrations. However, as discussed in Section 3 the presence of correlation between hourly average concentration observed in different hours and dependence of concentrations on time of day or period of the year can strongly affect this relationship. Air pollutants commonly show this correlation and time dependence and these effects must, therefore, be taken into account in developing suitable P_C functions. Again, as discussed in Section 3, taking these effects into account may make use of the case in which independence of hours and no time dependence of hours are assumed. This case will be discussed first.

If no correlation or time dependence of hours exists, Eq. (3-22) applies, namely:

$$P_C = (1 - G(C))^n \tag{4-8}$$

where P_C is the distribution of the highest concentration for n hours. (n= 8760 hrs. for one year or 43,800 hrs. for 5 years.) The function G(C) is defined by Eq. (3-4):

$$G(C) = Pr\,[C_{obs} > C] \tag{4-9}$$

That is, G(C) is the probability that an observed hourly average concentration is greater than C.

If the broader definition of P_C is used, Eq. (3-26) applies, where G(C) is substituted for p:

$$P_C^{(m)} = \sum_{\nu=0}^{m-1} \frac{n!}{\nu!(n-\nu)!} G^{\nu} (1 - G)^{n-\nu} \tag{4-10}$$

where $p_C^{(m)}$ is the distribution of the mth highest hourly average concentration for n hours.

Thus, if the distribution function G(C) is known, the desired P_C function can be obtained from application of Eq. (4-8) or (4-10). Studies have found that the distributions of short-term time averaged concentrations of air pollutants can usually be represented by lognormal, Weibull, or gamma distribution functions[5]. Of these the Weibull function provides a good fit to photochemical oxidant air monitoring data[6] and is convenient to use since its G(C) function can be stated explicitly as in Eq. (3-6).

$$G(C) = e^{-(C/\delta)^k} \tag{4-11}$$

The parameter δ is referred to as the scale factor. It is the concentration corresponding to G(C) = 0.368. It establishes the approximate position of the mid-concentration values of the distribution. The parameter k is called the shape factor. It tends to be a measure of the spread of the distribution. The larger k the more compact the distribution. If the values of k and δ have been determined for a given geographic area the corresponding P_C functions can then be obtained by use of Eqs. (4-8) and (4-11) or Eqs. (4-10) and (4-11) at the given level of air quality.

For the risk assessment it is necessary to connect alternative levels of the ambient air quality standard with the corresponding P_C function. This is easily done through the Weibull distribution, Eq. (4-11). The proposed form for the ozone standard is:

> C_{STD} ppm hourly average concentration with an expected number of exceedances per year less than or equal to E.

It is shown by Eq. (3-12) that for any region to which the Weibull function applies and just meets the standard:

$$G(C) = e^{-(\ln (n_E/E))\,(C/C_{STD})^k} \tag{4-12}$$

where:

C_{STD} = level of ambient air standard.

E = expected number of exceedances in n_E hours.

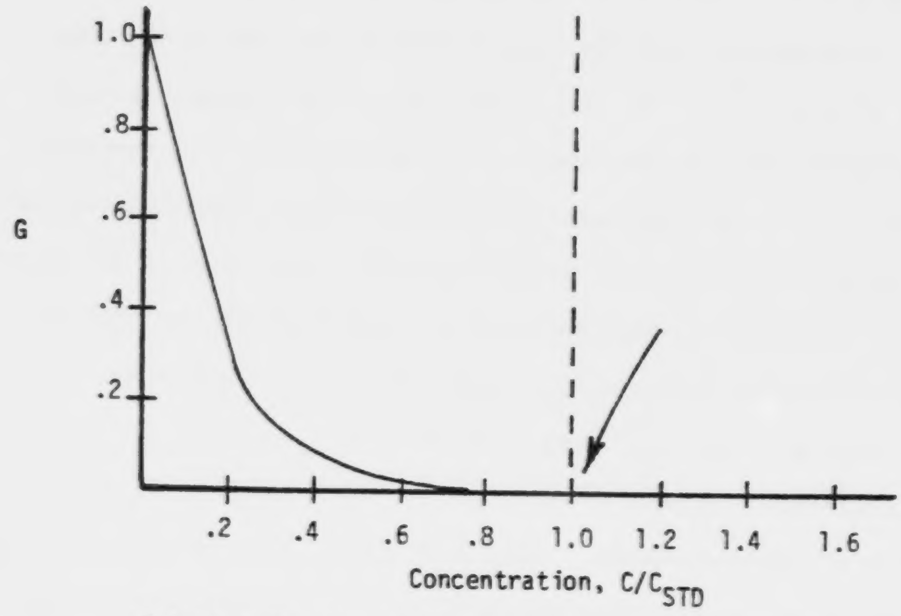
The term n_E is customarily the number of hours in one year or 8760 hours. The expected exceedance rate would normally be one for an air standard. In this case Eq. (4-12) becomes:

$$G(C) = e^{-9.078\,(C/C_{STD})^k} \tag{4-13}$$

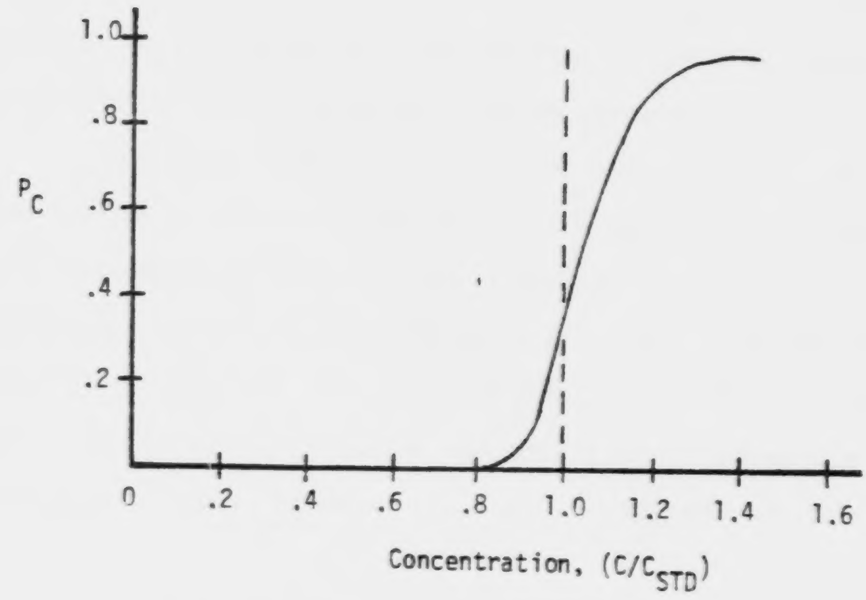
It should be pointed out that from the point of view of the risk assessment method developed in this report, the designation of the expected number of exceedances is a relatively arbitrary matter and could be set at any value that gives a convenient level for C_{STD} so long as the risk is the same (the value of m in $P_C^{(m)}$ of Eq. (4-10) has a more direct impact on health since it directly bears on the number of exceedances of a true threshold). For example, if it is decided that it is undesirable to have any exceedances of a true threshold over a given time period then the P_C used in the calculation is $P_C^{(1)}$ (see Eq. (4-10)) and once an acceptable level of risk is chosen any combination of E and C_{STD} values which yields this risk value in a given area give the same level of protection.

The relationship between the G(C) and P_C distribution for a Weibull distribution of hourly averages is shown in Fig. 4-3. The combination C_{STD} and E determine the general location of P_C while k determines its spread. Figure 4-4 shows the P_C functions for a one year period for a series of alternate levels of C_{STD} with E = 1 and for the Weibull shape factor k = 1. Figure 4-5 shows the effect of changing the shape factor at C_{STD} = 0.1 ppm and E = 1. Figure 4-6 shows the effect of changing k for a P_C covering a 5 year period. It is seen form the three figures that changing C_{STD} displaces the P_C function over a wide range while having a relatively small effect on its shape. Changing k causes little actual displacement of the P_C for a one year period but has a large effect on its shape. The changing k causes the P_C function in Fig. 4-5 to pivot about the point (0.10, 0.368). The effect of k on the P_C for a five year period (Fig. 4-6) is still largely in the shape of the function, but there also seems to be more displacement.

Figure 4-3 Hypothetical Distribution of Hourly Average Concentration and the Corresponding Distribution of the Annual Maximum Hourly Average Concentration



(a) Distribution of Hourly Average Concentration



(b) Distribution of Annual Maximum Hourly Average Concentration

Figure 4-4 Change P_C Function with Change in Standard Concentration Level

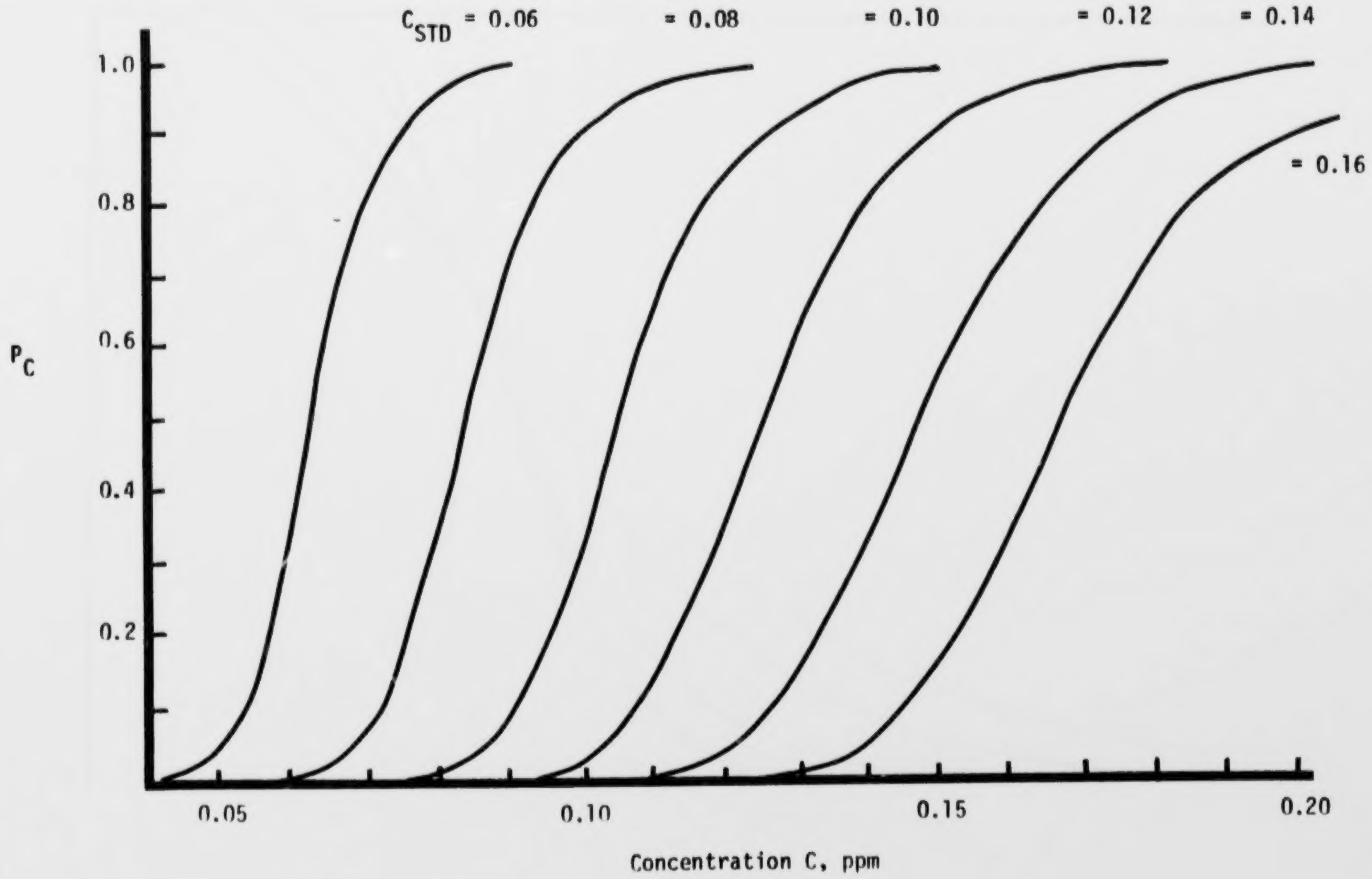


Figure 4-5 P_C Function for One Year Period for Different Values of Weibull Scale Factor, k

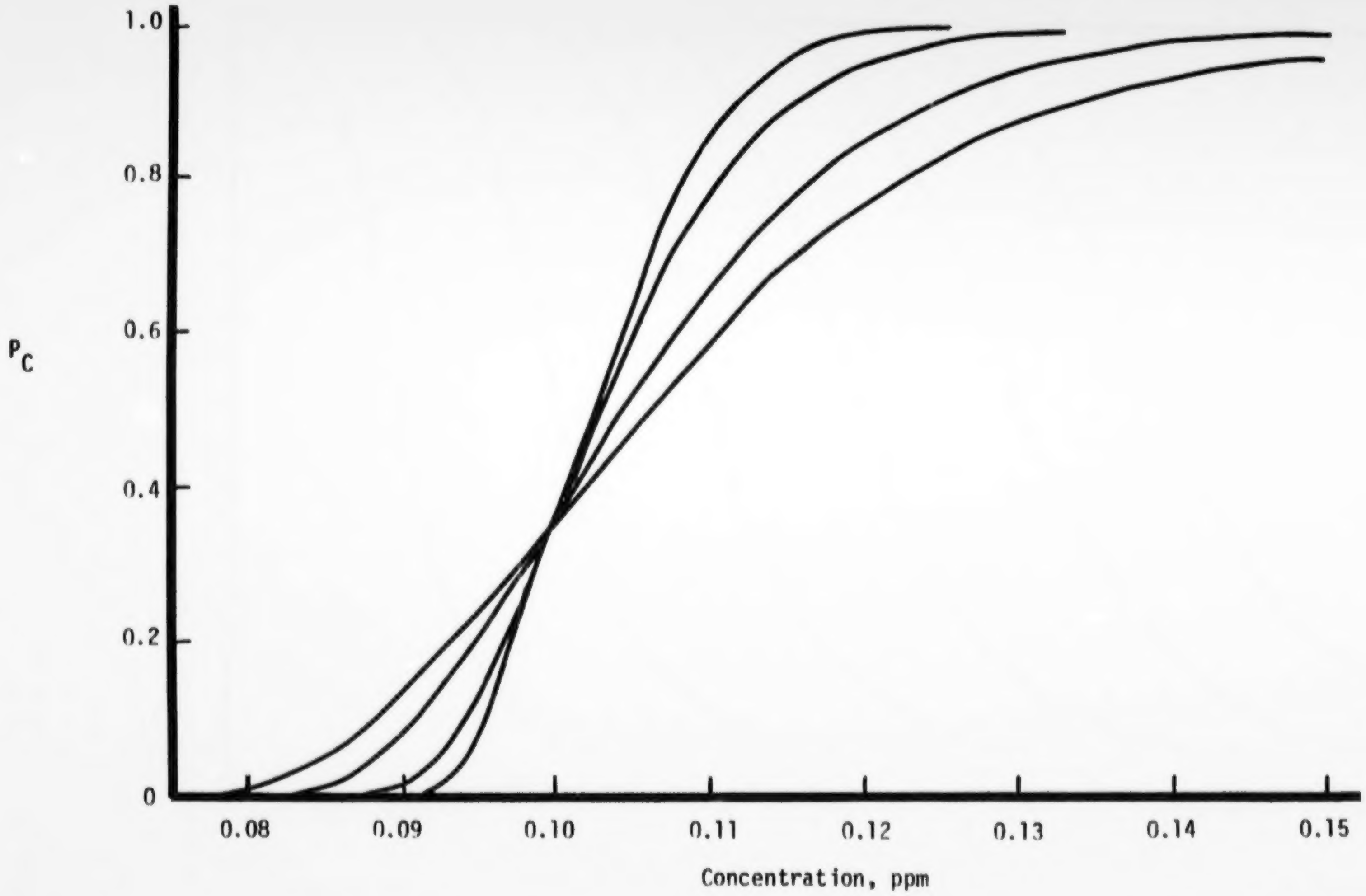
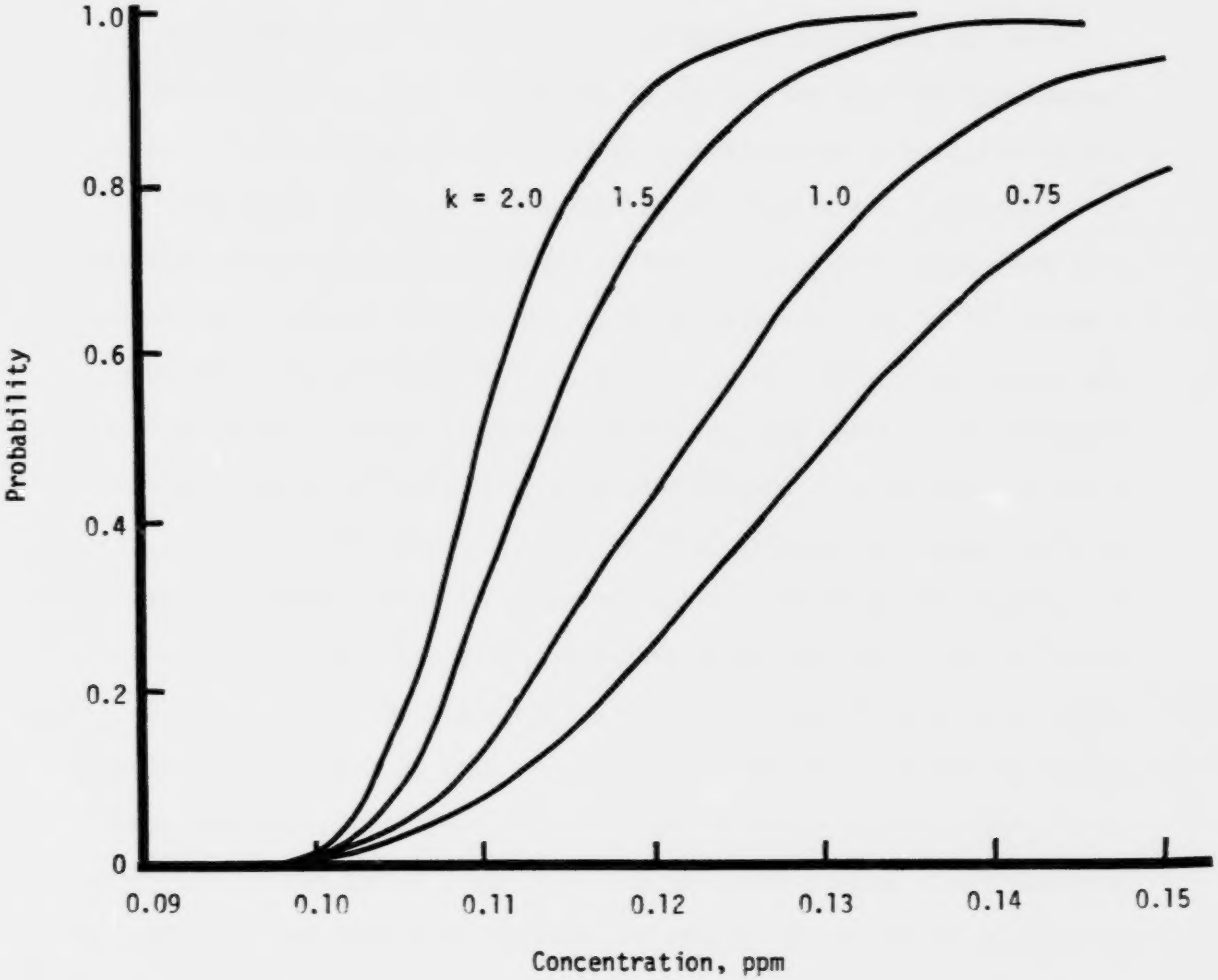


Figure 4-6 P_C Function for 5 Year Period for Different Values of Weibull Scale Factor, k



This results from the fact that the pivot point for the distribution is now very close to the concentration axis (0.10, 0.007). In general, calculated risk values will be more sensitive to changing values of C_{STD} than to changes in k over the usual ranges of these parameters.

From the preceeding discussion it is seen that the assumption of independence of hours and the use of the Weibull function to represent the distribution of hourly average concentrations readily yield P_C and $P_C^{(m)}$ functions. The Weibull can be used with little loss in accuracy even where other distributions such as the lognormal distribution provides a better fit to the concentration data. The primary concern in estimating the appropriate P_C function is to have it placed properly along the con-concentration axis and have the correct degree of spread. The parameters δ and k in the Weibull function provide wide flexibility in this regard. As shown above, the standard level essentially places the P_C function. The appropriate values of k can be obtained by fitting a Weibull distribution to hourly average concentration data obtained from air monitoring sites. The range of applicable k values for standard setting purposes should be determined by examining aerometric data in areas that are close to the concentration ranges of the alternate levels of the standard under consideration. Where this is not possible, the k values can be determined at existing levels of air quality and the results extrapolated to potential standard levels.

While using a Weibull distribution where a lognormal or gamma distribution might be more appropriate does not appear to lead to serious errors in the risk estimates, ignoring the possible dependence of hourly average concentrations can lead to significant error. As discussed in Section 3, internal EPA study[7] showed that dependence of one hourly concentration on the value of another did not lead to serious errors if independence of hours was assumed. It also showed that dependence of concentrations on time of day or year can lead to P_C functions which were placed lower on the concentration axis than would be obtained assuming no time dependence. When this time dependence can be modeled it should be possible to generate the corresponding P_C tables of functions and express them mathematically. This was done in the EPA study for the daily maximum hourly average ozone concentration.

Another approach can be taken as discussed in Section 3, if the time dependence is such that the maximum concentration tends to occur only within some determinable period and the probability distribution of concentrations is approximately the same for all hours within the time period. In this case all hours outside the time period can be excluded and independence assumed for the hours within the period. A Weibull distribution then could be fit to the hours within the time period to determine the appropriate k values and Eqs. (4-8) and (4-12) used to calculate P_C. To the extent that the period under consideration is also likely to contain the mth highest hourly average concentration, $P_C^{(m)}$ could be obtained using Eq. (4-10). The term n_E in Eq. (4-12) would be set equal to the hours in one calendar year of the time period. The term n in Eqs. (4-8) and (4-12) would be n_E times the number of years for which the risk is to be estimated. In the following section, this procedure is applied to ozone.

4.5 Determination of P_C Functions for Ozone

As suggested in the preceeding sections estimating the appropriate P_C function involves: 1. determination of mathematical function or functions which best describe the distribution of time averaged ambient concentrations of the pollutant; 2. given the distribution of time averaged concentration, the standard, and the nature of the correlation between concentrations in neighboring time periods and dependence upon time of day and year, derive sutiable P_C functions and 3. estimate the range of values of parameters appearing in the P_C function.

A study performed under contract for EPA involving 14 sites scattered around the United States and involving 22 site years of data showed that the Weibull distribution provides an excellent fit to hourly ozone concentrations[6]. In only two cases did a lognormal distribution give a superior fit.

Ozone hourly average concentrations exhibit strong correlation and strong dependence on time of day and time of year. The day of the week also has some effect. As indicated in Section 4-3, the correlation between neighboring concentrations does not appear to have an important effect on the P_C function; however, the time dependence does.

The method of dealing with the time dependence, as discussed, was to find a time period in which the time dependence was relatively constant and which was highly likely to contain the maximum hourly average concentration for the time period. If such a time period existed and a distribution could be fit to the time-averaged concentrations within this period, then the appropriate P_C function could be derived by assuming complete independence of the time averaged concentrations. For ozone this period is during the midday hours of the warm months of the year for urban areas.

In a continuation of the above mention study[6] it was found that Weibull distributions could be fit to data obtained between 11 AM and 6 PM both from May through September and July through August. It was further shown the maximum ozone concentration had significantly more chance of occurring during the longer of the two periods. Therefore, this time period, which contains 1071 hours, was used in the derivation of the P_C function.

The form of the National Ambient Air Quality Standard proposed for ozone is: C_{STD} ppm hourly average concentration with an expected number of exceedances per year less than or equal to one. For a region whose air quality just meets this standard the Weibull distribution of hourly averages for the hours 11 AM to 6 PM, May through September, would be according to Eq. (4-12):

$$G(C) = e^{-(\ln 1071)(C/C_{STD})^k} \tag{4-14}$$

And the P_C function for a period of n_y years would be from Eq. (4-8).

$$P_C = (1-e^{-(\ln 1071)(C/C_{STD})^k})^{1071\, n_y} \tag{4-15}$$

By substituting alternate levels, C_{STD}, of the standard into Eq. (4-15) the P_C function for n_y years needed for the risk assessment can be obtained. However, before this can be done it is necessary to estimate the range of values of the parameter k (Weibull shape factor) over different regions in the U.S. Table 3-11 shows measured k values for the 1071 hour time period as well as other time periods. The range for the values based on 1071 hours is 1.31 to 2.04.

For the risk assessment best estimates were made of the lower bound and upper bound values of k. This was done by probability encoding the contractor and a PEDCo researcher involved in the development of the Weibull distributions and the P_C functions. The information base was the data in Table 3-11, plus the data developed during the Weibull distribution studies. The data shown in Tabel 3-11 are, by and large, for geographic regions above the range of alternative ozone standards considered. The Weibull studies suggest that the k factors at the standard levels would be somewhat higher than those shown in Table 3-11. This factor was taken into account in the encoding. The contractor first encoded Mr. T. Johnson of PEDCo and then himself. The results of the encoding are shown in Tables 4-1 and 4-2. In encoding Mr. Johnson the upper and lower bounds of the variable were first encoded followed by the use of the interval method[3]. The subject was then allowed to make slight adjustments based on viewing the plotted probability function. When the contractor was his own subject the same approach was used but the final adjustments were made directly to the plotted curve. In this way a more completely defined curve was obtained.

The distributions were combined by Mr. T. Feagans of EPA into a single distribution for the lower bound and a single distribution for the upper bound (Table 4-3). The median values for the location of the lower and upper bound shape factors were 1.36 and 2.54 respectively. Since the range of k values varies somewhat with the standard level, the range, strictly speaking, should be estimated for each alternative level of the standard. In the case of ozone the difference is not likely to be large enough to seriously affect the risk estimates.

Table 4-1 Subjective Probability Distributions of Lower and Upper Bounds of Weibull Shape Factor k For n at Approximately 1000 Hours and Maximum Concentration at Approximately 0.1 ppm

Subject: Ted Johnson
PEDCo

Interviewer: W.F. Biller
Date: 1/26/78

Lower Bound k		Upper Bound k	
k	Cum. Probability	k	Cum. Probability
1.30	0.001	2.30	0.001
1.38	0.25	2.37	0.25
1.43	0.50	2.50	0.50
1.48	0.75	2.70	0.75
1.50	0.999	3.00	0.999

Table 4-2 Subjective Probability Distributions of Lower and Upper Bounds of Weibull Shape Factor k For n at Approximately 1000 hours and Maximum Concentration at Approximately 0.1 ppm

Subject: W.F. Biller

Interviewer: W.F. Biller
Date: 1/31/78

Lower Bound k		Upper Bound k	
k	Cum. Probability	k	Cum. Probability
0.7	0.01	2.3	0.01
0.8	0.02	2.4	0.09
0.9	0.05	2.5	0.30
1.0	0.10	2.6	0.60
1.1	0.16	2.7	0.75
1.2	0.27	2.8	0.83
1.3	0.60	2.9	0.88
1.4	0.86	3.0	0.91
1.5	0.95	3.1	0.94
1.6	0.99	3.2	0.96
		3.3	0.97
		3.4	0.98
		3.5	0.99

Table 4-3 Subjective Probability Distributions for Upper and Lower Bound Weibull Shape Factors (k) for Distributions of Hourly Average Ozone Concentrations

Probability that k is below specified value	Lower Bound k	Upper Bound k
0.10	1.15	2.41
0.30	1.30	2.50
0.50	1.36	2.54
0.60	1.41	2.66
0.90	1.47	2.98

4.6 Example Application of Method

Suppose that three health effects have been identified for a given pollutant and that there is some uncertainty as to the precise location of each of the effects. Suppose further that the evidence for one of the effects is very weak at the present time and there is considerable doubt the effect exists. However, if it does exist, it is a serious effect and, therefore, it should have some bearing on the setting of the standard. Suppose there is also some doubt about the existence of one of the other effects, but in this case the effect is not considered to be too serious. However, it also should be taken into account. It will be assumed there is no question about the existence of the third effect which is considered to be serious. A thorough review of the information available on each health effect by a group of medical experts establishes that the uncertainties in the locations of the effects are best described by normal distributions. The findings of the group are summarized as follows:

Health Effect	Most Probable Location of Threshold (ppm)	Standard Deviation of Density Function (ppm)	Probability Effect is Real (%)	Seriousness of Effect
1	0.15	0.05	15	Serious
2	0.20	0.04	80	Mild
3	0.30	0.08	100	Serious

In a real situation much more information would be supplied concerning the seriousness of the effect. It also would not be possible to describe the probability density function for the location of the threshold so easily. A tabulation of the function versus concentration for each effect would have to be developed.

Simple inspection of the table does not provide a clear indication as to where the level of the national ambient air standard should be set to provide an adequate margin of safety against all three effects simultaneously. However, application of the procedure outlined in the previous section gives an assessment of the medical risks involved at each potential level of the standard and can be used, therefore, in judging where the standard can be set to give an adequate margin of safety.

The margin of safety requirement means that an exceedance of the standard concentration will not, if the margin is adequate, generally result in an exceedance of a true health effect threshold. To apply the approach it is first necessary to adopt a form of the standard that provides protection appropriate to the medical effects under consideration. Suppose the effects are such that any exposure for a short period (such as one hour) over the thresholds are to be avoided. In this case one appropriate form would be:

> C_{STD} ppm hourly average with expected exceedances per year
> less than or equal to one.

This form is close to that of the current oxidant standard but differs from that form in that its statistical nature allows it to be directly related to yearly distributions of hourly average concentrations.

It is assumed that a Weibull distribution with $k = 1$ fits the ambient concentrations and there is no time dependence (i.e., $n_E = 8760$ hrs.). Also assume the risk is to be calculated for a one year period. It follows from the discussion in Section 4.3 that the P_C function is:

$$P_C = \left(1-e^{-(\ln 8760)(C/C_{STD})}\right)^{8760} \qquad (4\text{-}16)$$

Equation (4-16) plus the information supplied in the above table on the health effects thresholds can then be used to calculate for each level of the standard: 1. The risk, R, that no hourly average concentration will exceed any of the true health effect thresholds in a given year and 2. The individual risks, R_i, that no hourly average concentration will exceed the i-th true threshold in a given year. For the three thresholds the risk $R_{i,j}$ that no hourly average concentration exceeds either the i-th or j-th true threshold can also be calculated. The calculation of these risks for each standard level is achieved by:

1. Developing the appropriate P_C function for each level at which the standard may be set, in this case Eq.(4-16).
2. Forming the composite distribution function p_T° (C) for all three thresholds from Eq.(4-5).
3. Determining R by numerically integrating the product of $P_C(C)$ and p_T° (C) as indicated by Eq.(4-4).
4. Determining each R_i by carrying out the process in Eq.(4-4) for $P_C(C)$ and p_{T_i} (normal distributions in this example).
5. Repeating #3 and #4 for any composite of two threshold functions.

When this procedure is carried out for the above three health effects the tabulation shown in Table 4-3 is obtained (Details of the calculations are given in Appendix D).

It is seen that if the standard is set at 0.10 ppm, the margin of safety is such that there is a 0.05 risk (or one chance in 20) that a health effect threshold will be exceeded in a given year. At the same

Table 4-4 Risks of an Hourly Average Concentration Exceeding True Health Effect Thresholds in a Given Year for Different Values of a National Ambient Air Quality Standard

1	2	3	4	5	6
Standard (ppm)	R_{MT}	R_1	R_2	R_3	$R_{1,3}$
0.06	0.009	0.007	0.001	0.002	0.008
0.07	0.013	0.010	0.001	0.002	0.013
0.08	0.022	0.015	0.002	0.004	0.019
0.09	0.033	0.022	0.005	0.006	0.028
0.10	0.050	0.030	0.011	0.008	0.038
0.11	0.072	0.040	0.022	0.012	0.052
0.12	0.102	0.050	0.040	0.017	0.067
0.13	0.141	0.062	0.066	0.025	0.085
0.14	0.190	0.073	0.101	0.034	0.104
0.15	0.248	0.085	0.147	0.045	0.126
0.16	0.315	0.096	0.202	0.060	0.149
0.17	0.383	0.106	0.265	0.077	0.175
0.18	0.457	0.115	0.334	0.098	0.202

time the risks of exceeding the individual thresholds are 0.03 (1 in 33), 0.011 (1 in 91), and 0.008 (1 in 125). Since the first and third effects are serious, this level of protection might be considered as the minimal acceptable value by the EPA decision makers. If so, the standrard would be set at 0.10 ppm. Considering, however, that the second effect is mild, it is of interest to determine the probability of not exceeding either the first or third threshold. This probability is shown in Column 6 of the table. If the .05% level is still considered acceptable the standard could be set at 0.11. This level would still give a 96.0% probability of not exceeding the mild health effect.

The above example is greatly simplified since it is primarily intended to show how the equations developed in the previous section are used. In actual application of the method other considerations would enter. Furthermore, considerable power can be added to the method through a sensitivity analysis of the uncertainties involved in the determination of the P_C and p_T functions. These added considerations are treated elsewhere[4].

4.7 Conclusions and Recommendations

It is concluded that the method of risk assessment presented in this section is a practical method for developing national ambient air quality standards which are in accord with the intent and spirit of the Clean Air Act as presently written. The method leads naturally to the use of standards expressed in statistical forms. It gives a proper perspective to viewing exceedances of a standard level. The real concern is to protect against pollutant concentrations exceeding health effects thresholds and to set the standard at a level which reduces the risk of such exceedances to an acceptable level. Exceedances of the standard level do not necessarily mean a health effects threshold has been exceeded and the risk that they do can be set as low as the EPA Administrator deems necessary to protect the public health.

The risk assessment method also is a logical starting point for:
1. estimating the expected number of people in a population who will suffer health effects at given air quality levels and the uncertainties in such estimates and 2. Cost-benefit analyses.

In those aspects of the method to which the contractor contributed there is a need to: determine P_C functions accurately, determine the extent to which more accurate functions can be developed, and to obtain more complete information on the range of P_C functions experienced in different parts of the United States. In investigating the question of improved P_C functions it will be necessary to further study the interdependence of time averaged concentrations and the nonstationarity of the underlying stochastic processes which are responsible for observed pollutant levels and to determine how these factors affect the P_C functions.

There is also a need to better define the behavior of aerometric data and, therefore, the P_C functions at the standard level.

It is recommended that the above mentioned problem areas be further investigated. In addition it would be worthwhile to extend the basic method to estimating expected numbers of people affected at air pollution levels above the health effects thresholds and the attendant uncertainties and risks.

4.8 References

1. Clean Air Act, Section 109, 42 U.S.C.

2. "A Legislative History of the Clean Air Amendments of 1970", U.S. Senate Committee on Public Works, 1970.

3. C.S. Spetzler and C.A.S. Stael von Holstein, "Probability Encoding in Decision Analysis", Management Science, Vol. 22, No. 3, November 1975.

4. "A Method for Assessing Health Risks Associated With Alternative Air Quality Standards for Photochemical Oxidants", Report by U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, in press.

5. T.C. Curran and N.H. Frank, "Assessing the Validity of the Lognormal Model When Predicting Maximum Air Pollution Concentrations", Presented at 68th Annual Meeting of the Air Pollution Control Association, Boston, June 15-20, 1975.

6. "The Validity of the Weibull Distribution as a Model for the Analysis of Ambient Ozone Data". Draft Report to EPA by PEDCO Environmental, November 17, 1977.

7. Private Communication, J. Horowitz, Office of Air and Waste Management, U.S. Environmental Protection Agency, Washington, D.C.

APPENDIX A
SUPPORT MATERIAL FOR SECTION 3

1. Derivation of Approximate Relationship (3-18)

Starting with Equation (3-16):

$$P_m = \frac{n!}{m!(n-m!)} \left(\frac{E}{n_E}\right)^m \left(1 - \frac{E}{n_E}\right)^{n-m} \qquad \text{(A-1)}$$

The following approximately holds for n >> m:

$$\frac{n!}{(n-m!)} \approx n^m \qquad \text{(A-2}$$

If define a constant λ such that:

$$\lambda = \frac{n}{n_E} \qquad \text{(A-3)}$$

Eq. (A-1) is approximated by:

$$P_m \approx \lambda^m E^m \left(1 - \frac{E}{n_E}\right)^{n-m} \qquad \text{(A-4)}$$

If $E << n_E$ the term in brackets can be expanded as a binomial series to yield:

$$\left(1 - \frac{E}{n_E}\right)^{n-m} \begin{aligned} &= 1 - \left(\frac{n-m}{n_E}\right)E + \frac{(n-m)(n-m-1)}{2!n_E} E^2 + \cdots \\ &\approx 1 - \lambda E + \frac{\lambda^2 E^2}{2!} - \frac{\lambda^3 E^3}{3!} + \cdots\cdot \end{aligned}$$

The approximation series expression is the expansion of the exponential function. Therefore:

$$(1-\frac{E}{n_E})^{n-m} \simeq e^{-\lambda E}$$

Substituting into the approximation (A-4) yields:

$$P_m \simeq \frac{\lambda E}{m!} e^{-\lambda E} \tag{A-5}$$

which is Expression (3-18) appearing in Section 3.

2. Demonstration That the Concentration Corresponding to the Most Probable Value of the mth Highest Concentration in n Hours is the Concentration With an Expected Exceedance Rate of m.

This relationship is true if the distribution of hourly average concentrations is described by a Weibull function with $k = 1$. To find the concentration corresponding to the most probable value of the mth highest concentration in n hours the derivative of Eq. (3-27) is set equal to zero. To simplify the expression use the δ notation for the Weibull and set $k = 1$. Thus:

$$f(C_{(m)}) = \left[\frac{1}{\delta} \frac{n!}{(m-1)!(n-m)!} \right] (1-e^{-C/\delta})^m \tag{A-6}$$

If the derivative of Eq. (A-6) is taken with respect to C and all terms cleared that are not zero at the maximum value, the following expression is obtained.

$$\frac{m}{n} = e^{-C^\circ/\delta} \qquad \text{(A-7)}$$

Equation (A-7) has the same form as Eq. (3-8):

$$\frac{E}{n_E} = e^{-C_E/\delta} \qquad \text{(A-8)}$$

Therefore:

$$C^\circ = C_E \qquad \text{(A-9)}$$

$$E = m \qquad \text{(A-10)}$$

3. Demonstration of Expression (3-37)

Using the following series expansion:

$$(1 - X)^n = 1 - nX + \frac{n(n-1)}{2!} X^2 - \cdots\cdots$$

the right hand side of Eq. (3-35) expands to:

$$(1 - \frac{n_o}{n} p)^n = 1 - n_o p + \frac{(n-1)^{n_o^2}}{2!} p^2 + \cdots\cdots$$

$$\simeq 1 - n_o p \quad \text{for p very small}$$

Similarily the right side of Eq. (3-36) also becomes approximately $(1 - n_o p)$ for small p. Since we are concerned with the highest observations, the value of p is much less than one and the approximation should hold closely. Note also, this result is independent of the value k.

APPENDIX B

DERIVATION OF BASIC EQUATIONS FOR ASSESSING HEALTH RISKS ASSOCIATED WITH ALTERNATIVE AIR QUALITY STANDARDS

1.0 Risk of Exceeding a True Health Effect Threshold or the Lowest of Several Thresholds

First determine probability P where:

P = Probability that no hourly average concentration exceeds the true health effect threshold or the lowest of several thresholds in a given period.

Let:

$P_C(C)$ = Probability that no time averaged concentration exceeds the concentration C in the given period.

$p_T(C)$ = Probability density function expressing the uncertainty in the location of the true threshold or the lowest of several thresholds.

Note:

$\int_a^b p_T(C)dC$ = Probability that a true threshold or the lowest of several thresholds is in the interval a,b.

It follows that:

$P_C(C)p_T(C)dC$ = Probability that the true threshold or the lowest of several thresholds is contained in the interval dC and this value is not exceeded by any hourly average concentration in the given period.

By integrating over all values of C the quantity P, defined above, is determined:

$$P = \int_0^\infty P_C(C)P_T(C)dC \tag{B-1}$$

If R is the probability or risk of exceeding a health effects threshold or the lowest of several thresholds one or more times in the given period, then:

$$R = 1 - P$$

$$R = 1 - \int_0^\infty P_C(C)p_T(C)dC \tag{B-2}$$

Note that if the function P_C is defined as the probability that no time averaged concentration exceeds the concentration C it must also be the probability that the highest time averaged concentration observed in the time period does not exceed C (therefore, is $\leq$ C). In other words, P_C is the cumulative distribution of the highest time averaged concentration occurring in the given time period.

It is further noted that if it is desired to estimate the risk of exceeding a health effects threshold or the lowest of several threshold "m" or more times in a given time period, then P_C becomes the probability that a threshold will not be exceeded more than m-1 times in the given time period. In this case P_C is the probability distribution of the mth highest time averaged concentration.

2.0 Determination of Probability Density Function for the Lowest of Several Health Effect Threshold

Assume n different health effects and that a probability density function has been obtained for each. If the function $P_C(C)$ is known, Eq. (B-1) can then be used to evaluate the individual probabilities for each health effect.

$$P_i = \int_0^\infty P_C(C) P_{T_i}(C)dC \qquad \text{(B-3)}$$

Let:

P = The probability that no hourly average concentration exceed any of the n true thresholds in the given time period.

To evaluate this probability for all possible configurations it is assumed that each threshold in turn is the lowest threshold.

Differential elements of the following type can be formed:

$$P_C(C) P_{T_1}(C)dC \cdot \int_C^\infty P_{T_2}(x)dx \cdot \int_C^\infty P_{T_3}(x)dx \cdots \int_C^\infty P_{T_n}(x)dx$$

This term is the probability that: 1. no hourly average concentration exceeds the value C; 2. the threshold T_1 is in the interval dC; and 3. all the other thresholds are above this value.

If:

$$Q_i(C) = \int_C^\infty P_{T_i}(x)dx$$

And:

$$Q(C) = \prod_{i=1}^{n} Q_i(C)$$

The above expression simplifies to:

$$P_C \frac{Q}{Q_1} p_{T_1} \; dC$$

Integrating over the concentration range gives:

$$\int_0^\infty P_C \frac{Q}{Q_1} p_T \; dC$$

This integral is the probability that no hourly average concentration exceeds any of the n thresholds in the time period and T_1 is the lowest threshold. In order to enumerate all cases, it is necessary to form the sum of integrals of this type in which each threshold in turn is assumed to be the lowest.

Therefore:

$$P = \sum_{i=1}^{n} \int_0^\infty P \frac{Q}{Q_1} p_{T_i} \; dC$$

or:

$$P = \int_0^\infty P_C \left[Q \sum_{i=1}^{n} p_{T_i}/Q_i\right] dC$$

or:

$$P = \int_0^\infty P_C \, P_T \; dC$$

This equation is, as would be expected, identical to Eq. (B-1) except that p_T is now a composite probability density function such that:

$$p_T = Q \prod_{i=1}^{n} p_{T_i}/Qi \qquad \text{(B-4)}$$

Where:

$$Q_i(C) = \int_C^{\infty} p_{T_i} \, dC \qquad \text{(B-5)}$$

and

$$Q = \prod_{i=1}^{n} Q_i \qquad \text{(B-6)}$$

The quantity $p_T(C)dC$ is the probability that the lowest of the n thresholds is in the interval dC.

An alternate method of deriving the composite probability density function is to start with the composite cumulative distribution for the health effects threshold. Let:

$$D_T = Pr\ (T_{lowest} \leq C)$$

and:

$$D_{T_i} = Pr\ (T_i \leq C)$$

It follows from probability theory that:

$$1-D_T = \prod_{i=1}^{n} (1-D_{T_i}) \qquad \text{(B-7)}$$

The composite probability density function is by definition:

$$p_T = \frac{dD_T}{dC} \qquad \text{(B-8)}$$

Differentiating (B-7) yields:

$$\frac{dD_T}{dC} = \left(\prod_{i=n}^{n} (1-D_{T_i}) \right) \sum_{i=1}^{n} \frac{1}{1-D_{T_i}} \frac{dD_{T_i}}{dC} \qquad \text{(B-9)}$$

Since by definition:

$$1-D_{T_i} = \int_C^{\infty} p_{T_i}\, dC = Q_i \qquad \text{(B-10)}$$

$$Q = \prod_i (1-D_{T_i})$$

$$p_{T_i} = dD_{T_i}/dC \qquad \text{(B-11)}$$

Equation (B-9) is equivalent to Equation (B-4).

Equations (B-7) and (B-8) provide an alternate route to obtaining the composite density function, p_T, which can then be used in Eq. (B-2) to estimate risk.

3.0 Inclusion of Uncertainty that One or More Health Effects Exist

For some health effects there may be uncertainty that the effect actually occurs in humans. It would be desirable to include this consideration when considering the uncertainty in the location of the threshold on the concentration axis. Given uncertainty only in the location of the threshold it has been shown that Eq. (B-2) gives the risk that the true threshold T_i is exceeded one or more times in a given period.

If there is uncertainty as to whether the ith effect occurs in humans, assign the probability e_i that the effect does exist. Then choose a value of $C = u$ such that u is many times larger than any concentration likely to be encountered. In other words, u is many times beyond the concentration range of interest. In this case, if $p^{\circ}_{T_i}(C)$ is defined as the probability density function for the location of the ith effect if it does exist. A new overall probability density function can be written:

$$p_{T_i}(C) = e_i p^{\circ}_{T_i} + (i-e_i)\, \delta(C-u) \qquad i = 1 \cdots\cdots, n \qquad \text{(B-12)}$$

where $\delta(C-u)$ is the Dirac delta function. It has the property:

$$\int_a^b \delta(C-u)dC \quad \begin{cases} = 1 & \text{if } a \le u \le b \\ = 0 & \text{if } a > u \text{ or } b < u \end{cases} \qquad \text{(B-13)}$$

In other words, if the interval of integration includes $C = u$ the value of the integral is unity. If it does not include $C = u$ the integral is zero. The Dirac delta function itself may be considered as zero for all values of C except u. At u it has an infinite value. It is also assumed that $p^\circ_{T_i}(C)$ is essentially zero in the vicinity of $C = u$ and above.

Equation (B-12) is based on the premise that saying an effect does not exist is mathematically equivalent to saying that the true threshold is at a very high concentration which is above any concentration likely to be encountered. Thus, if the probability e can be assigned to the certainty that it does exist, the fraction e of the total area under the probability density curve can be assigned to the concentration range in which the effect is thought to be located if it does occur. The rest of the area, (1-e), can be assigned to a range above any concentration likely to occur.

Note also, the use of the Dirac delta function is only a matter of convenience since it leads to a simple form of Eq. (B-12). Any function can be used in this outer region as long as it has the value zero in the region of interest and has the area (1-e). The Dirac function is simply a convenient form for including the desired property in the probability density function.

Substituting Eq. (B-12) into (B-3) gives:

$$P_i = e_i \int_0^\infty P_C p^\circ_{T_i} dC + (1-e_i) \int_0^\infty P_C \, \delta(C-u)dC$$

Since P_C will be essentially 1 at $C = u$, the above equation yields:

$$P_i = e_i \int_0^\infty P_C p_{T_i} dC + (1-e_i) \qquad i = 1,\cdots, n \qquad \text{(B-14)}$$

To find the probability that no hourly average concentration exceeds any of n thresholds, it is necessary to derive the appropriate form of the composite threshold probability density function. Its general form is given by Eq. (B-4)

Substituting Eq. (B-12) into Eq. (B-4) and rearranging gives:

$$p_T = Q \sum_{i=1}^{n} \frac{e_i}{Q_i} p^{\circ}_{T_i} + \left[Q \sum_{i=1}^{n} \frac{1-e_i}{Q_i} \right] \delta(C-u) \qquad \text{(B-15)}$$

The first term on the right side of the Eq. (B-15) is evaluated in the ambient concentration range of the pollutant well below the value C = u. In this region the functions Q_i have a simple interpretation. Substituting Eq. (B-12) into Eq. (B-5).

$$Q_i = e_i \int_C^{\infty} p^{\circ}_{T_i} dC + (1-e_i) \int_C^{\infty} \delta(C-u) dC \qquad \text{(B-16)}$$

For all C < u

$$Q_i = e_i \int_C^{\infty} p^{\circ}_{T_i} dC + (1-e_i) \qquad \text{(B-17)}$$

The second term on the right side of the Eq. (B-15) is evaluated in the vicinity of C = u, far above the ambient concentration range. In this region the behavior of the Q_i functions needs to be more carefully considered. The problem is the behavior of the second term on the right side of Eq. (B-16) as C passes through u.

Note that in the vicinity of C = u the first term on the right side of Eq. (B-16) is zero by definition.

Thus:

$$Q_i = (1-e_i) \int_C^\infty \delta(C-u)dC \qquad \text{for } C \sim u \tag{B-18}$$

Expanding the second term of Eq. (B-15) gives:

$$[\, Q_2Q_3 \cdots\cdot Q_n(1-e_1) + Q_1Q_3 \cdots Q_n(1-e_2) + \cdots + Q_1Q_2 \cdots Q_{n-1}(1-e_n) \,]\, \delta(C-u)$$

which on substituting Eq. (B-18) yields:

$$\left\{ n\, \delta(C-u) \ [\int_C^\infty \delta(C-u)dC \ ]^{n-1} \right\} \prod_i (1-e_i)$$

The portion of this term in large brackets can be shown to have the properties of a Dirac delta function (Eq. (B-13)) as follows:

Let:

$$\psi(C-u) = n\, \delta(C-u) \ [\int_C^\infty \delta(C-u)dC \ ]^{n-1} \tag{B-19}$$

and:

$$\phi(C) = \int_C^\infty \delta(C-u)dC \tag{B-20}$$

then:

$$\frac{d\phi(C)}{dC} = -\delta(C-u) \tag{B-21}$$

Substituting into the bracketed terms:

$$\psi(C-u) = -\, n\phi^{n-1} \frac{d\phi}{dC}$$

or

$$\psi(C-u) = -\frac{d\phi^n}{dC} \tag{B-22}$$

To show that the derivative has the properties of the Dirac delta function, integrate between the limits a and b.

$$\int_a^b \psi(C-u)dC = -\int_a^b d\phi^n = \phi(a)^n - \phi(b)^n$$

$$= \left[\int_a^\infty \delta(C-u)dC\right]^n - \left[\int_b^\infty \delta(C-u)dC\right]^n \tag{B-23}$$

Using Eq. (B-13) to evaluate the quantities in brackets for different values of a and b it follows that:

$$\int_a^b \psi(C-u)dC \quad \begin{array}{l} = 1 \text{ if } a < u < b \\ \\ = 0 \text{ if } a > u \text{ or } b < u \end{array} \tag{B-24}$$

where b > a.

Therefore, $\psi(C-u)$ is also a Dirac delta function.

We can, therefore, write:

$$p_T = (Q \sum_i \frac{e_i}{Q_i} p^\circ_{T_i}) + \delta(C-u) \prod_i^n (1-e_i) \tag{B-25}$$

where:

$$Q_i = e_i \int_C^\infty p^\circ_{T_i} dC + (1-e_i)$$

or alternatively:

$$Q_i = 1-e_i \int_0^C p^\circ_{T_i} dC \tag{B-26}$$

and

$$Q = Q_1 Q_2 \cdot\cdot\cdot\cdot Q_n \tag{B-27}$$

Substituting Eq. (B-25) into Eq. (B-1) gives:

$$P = \int_0^\infty P_C p^\circ_T dC + \prod_i (1-e_i) \tag{B-28}$$

and therefore:

$$R = 1 - \int_0^\infty P_C p^\circ_T dC + \prod_i (1-e_i) \tag{B-29}$$

The Eq. (B-25) through (B-29) are the basic working equations for estimating risk when a multiple number of health effects thresholds are involved, and where one or more of the health effects may not occur in humans.

APPENDIX C

EXAMPLES OF COMPOSITE HEALTH EFFECT THRESHOLD DISTRIBUTIONS FORMED FROM SIMPLE HYPOTHETICAL DISTRIBUTIONS FOR INDIVIDUAL THRESHOLDS

The following simple examples will help illustrate the relationship of individual health effect threshold probability density functions and their composite functions. Assume two health effects with probability density distributions. The composite density function is:

$$p_T = e_1 Q_2 p_1 + e_2 Q_1 p_2 + (1-e_1)(1-e_2)\,\delta\,(C-u) \qquad \text{(C-1)}$$

(The superscripts on the p_i are omitted for convenience.)

Assume: 1. p_1 and p_2 are box-like functions; 2. p_2 lies beyond p_1 on the concentration axis with no overlap; 3. both effects are certain to exist. Eq. (C-1) then becomes:

$$p = Q_2 p_1^{\circ} + Q_1 p_2^{\circ} \qquad \text{(C-2)}$$

or

$$p = p_1 \int_C^{\infty} p_2 dC + p_2 \int_C^{\infty} p_1 dC \qquad \text{(C-3)}$$

This situation is shown in Figure C-1(a). The area under each box is 1.

In the interval [0,a], p_1 and p_2 are both 0, so that the composite distribution $p_T = 0$. In the interval [a,b], $p_1 = p_o$, $Q_2 = 1$, $p_2 = 0$. Therefore, $p_T = p_o$. Similarily it can be shown that above point b, $p_T = 0$. Thus, the plot of the composite interval (Fig. C-1(b)) is identical to the distribution p_1 alone.

This result is expected because in this example it is known with certainty that Threshold #1 is below Threshold #2. Thus, the probability that the lowest threshold is in a given interval is determined only by the density function for Threshold #1.

Figure C-1(c) shows the composite probability density function for a different situation in which the density functions for Threshold #1 and #2 are both box-like functions and are identical. That is, both have the same distribution as shown for p_1 in Fig. C-1(c). Since the density functions for the individual thresholds are uniform over the interval [a,b], the most probable point at which the lowest of the two will be encountered is at the lowest end of the interval [a,b]. If there are several identical, fully overlapping box-like threshold distributions, the composite distribution becomes curvilinear (Fig. C-1(d)) with more of its area displaced toward the lower end of the interval. The area under the composite function is, of course, one in both Figs. C-2(c) and (d).

Figure C-2(a) depicts a situation similar to that shown in Fig. C-1(a) in all respects except that there is only a 30% certainty that the first health effect exists. In this case the probability density function for Threshold #1 is divided between the region in which the threshold would be located if it does exist and a region of very high concentration above any concentration that could be encountered in the ambient atmosphere.

Figure C-1

Composite Threshold Probability Densities When Health Effects Are Certain to Exist

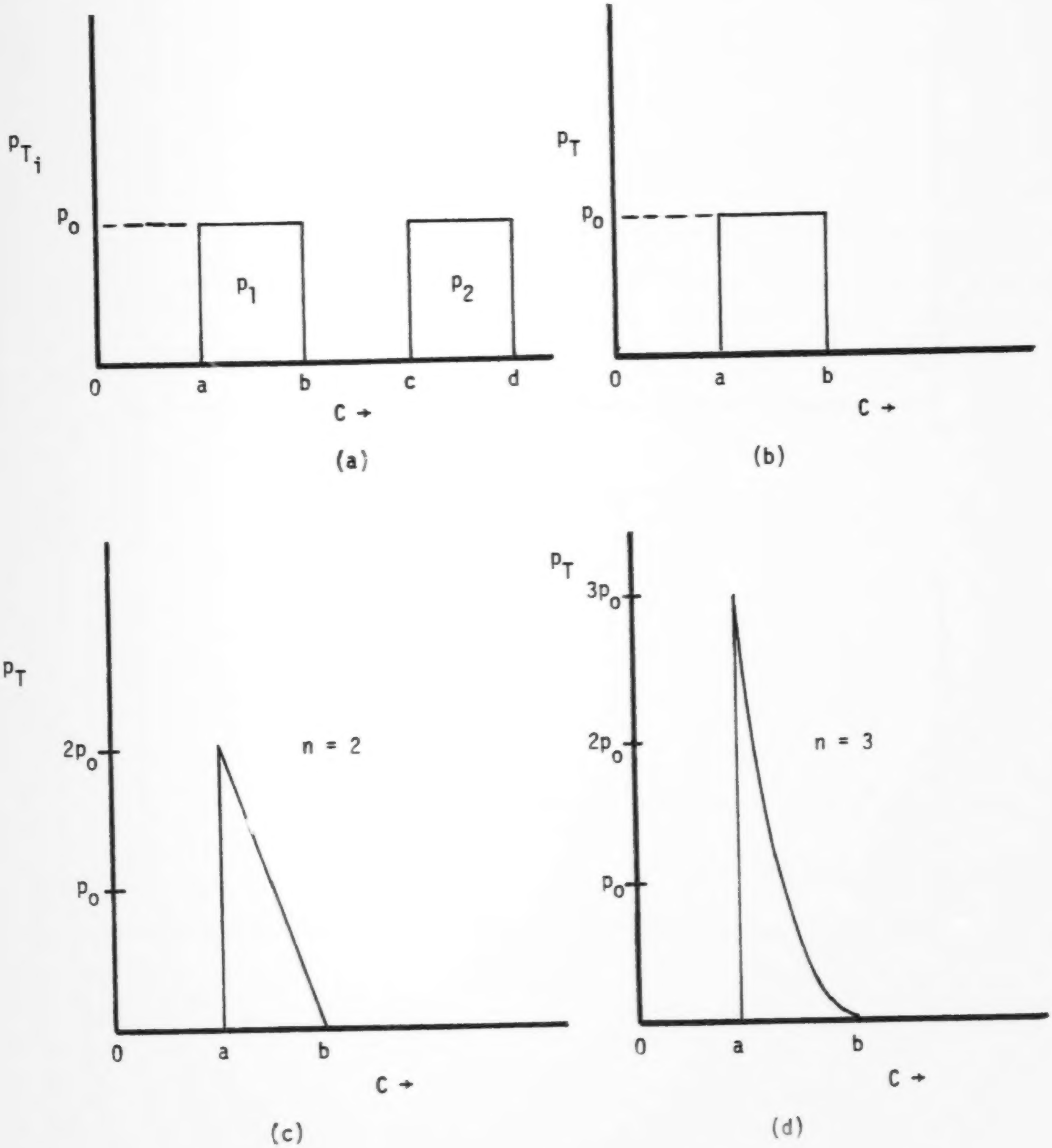
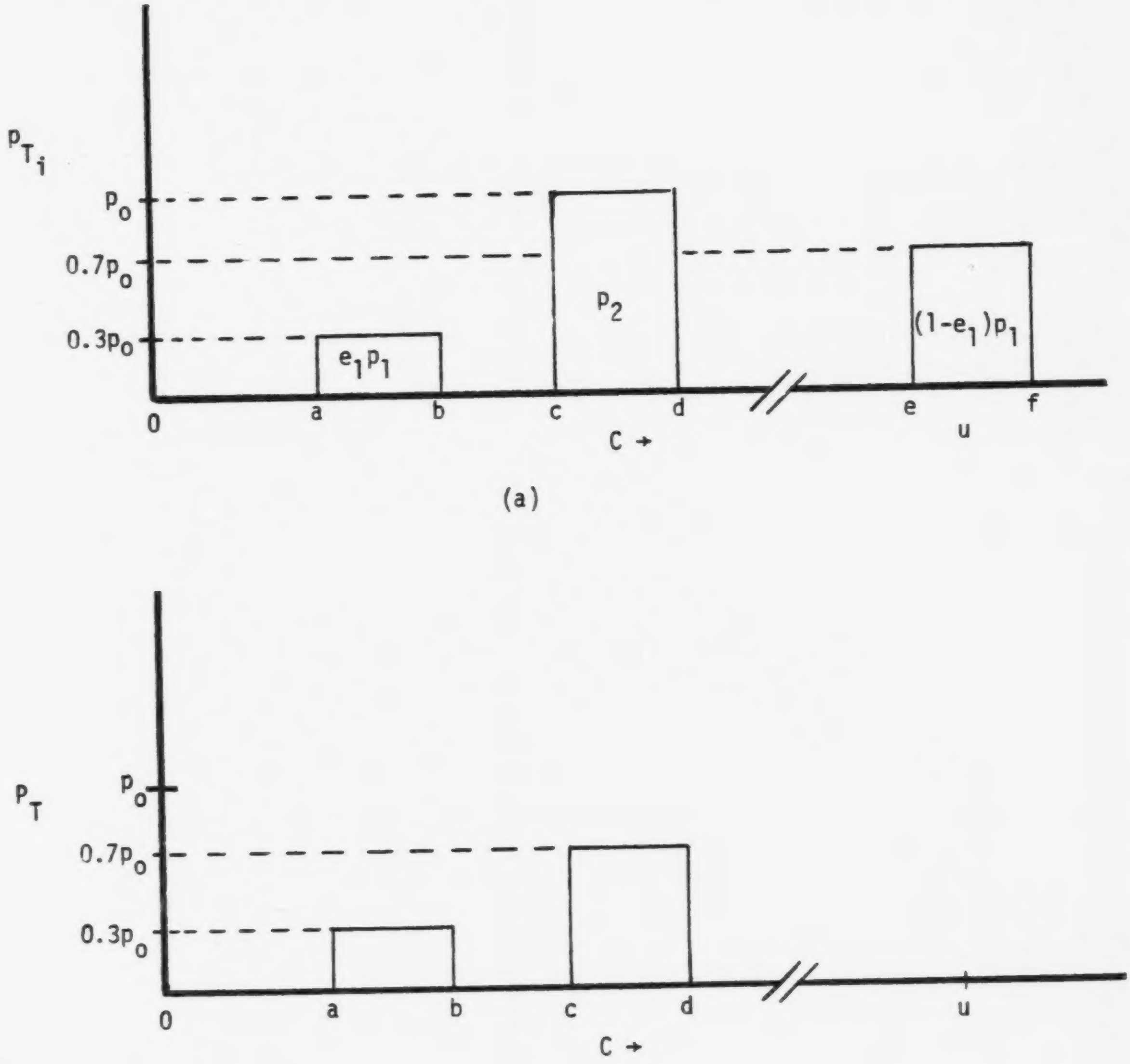


Figure C-2

Composite Threshold Probability Densities When the Existence of One Or More Health Effects is Not Certain



(a)

(b)

For graphical purposes a simple box function is shown in this region instead of the Dirac delta function. Both yield the same result.

The composite function can be written:

$$p_T = Q_2 e_1 p_1 + Q_1 p_2 \qquad \text{(C-4)}$$

where:

$$Q_1 = e_1 \int_C^\infty p_1 dC + (1-e_1) \qquad \text{(C-5)}$$

$$Q_2 = \int_C^\infty p_2 dC \qquad \text{(C-6)}$$

Note the absence of a third term in Eq. (C-4) which would give the value of p_T in the region of $C = u$. This term is zero because $e_2 = 1$ and, therefore, $(1-e_1)(1-e_2)\ \delta(C-u) = 0$. The graph of the composite function (Fig. C-2(b)) is obtained by evaluating (C-4) at each point on the concentration axis.

In the interval [0,a] both p_1 and p_2 are zero and, therefore, $p_T = 0$.

In the interval [a,b] $p_1 = p_o$, $p_2 = 0$. Thus, only the first term on the right side of the equation contributes. Since in this interval $Q_2 = 1$ and $e_1 = 0.3$, $p_{MT} = 0.3\ p_o$.

In the interval [b,c] again both p_1 and p_2 are zero and, therefore, $p_T = 0$.

In the interval [c,d] the p_1 term in Eq. (C-4) is zero, but the second term, $Q_i p_2$ has the value (1-0.3 x 1) x p_o = 0.7 p_o. Therefore, p_T = 0.7 p_o.

In the interval above point d, $p_1 = 0$, $p_2 = 0$ and the $\delta(u)$ term has a zero coefficient. Therefore, $p_T = 0$.

The overall area under the two boxes is 1, as expected.

It is seen that Fig. C-2(b) is very different from the Fig. C-1(b). The introduction of uncertainty that Threshold #1 exists leads in this case to a bimodal distribution.

When applying the above treatment to actual health effect problems it is, of course, not likely that simple box functions will be encountered. There is more likely to be considerable tailing of the distributions and overlapping. The box functions, nonetheless, serve to illustrate some important aspects of the relationship between the composite density function and its component threshold functions.

APPENDIX D

DETAILS OF CALCULATIONS PERFORMED IN SECTION 4.5

EXAMPLE APPLICATION OF METHOD

The table of risks in Section 4.5 was obtained as follows:

D.1 Develop the Appropriate P_C Function for Each Alternate Standard Level

Rearranging Eq. (4-16) of Section 4, the constant in the exponential term can be calculated for each level of the standard:

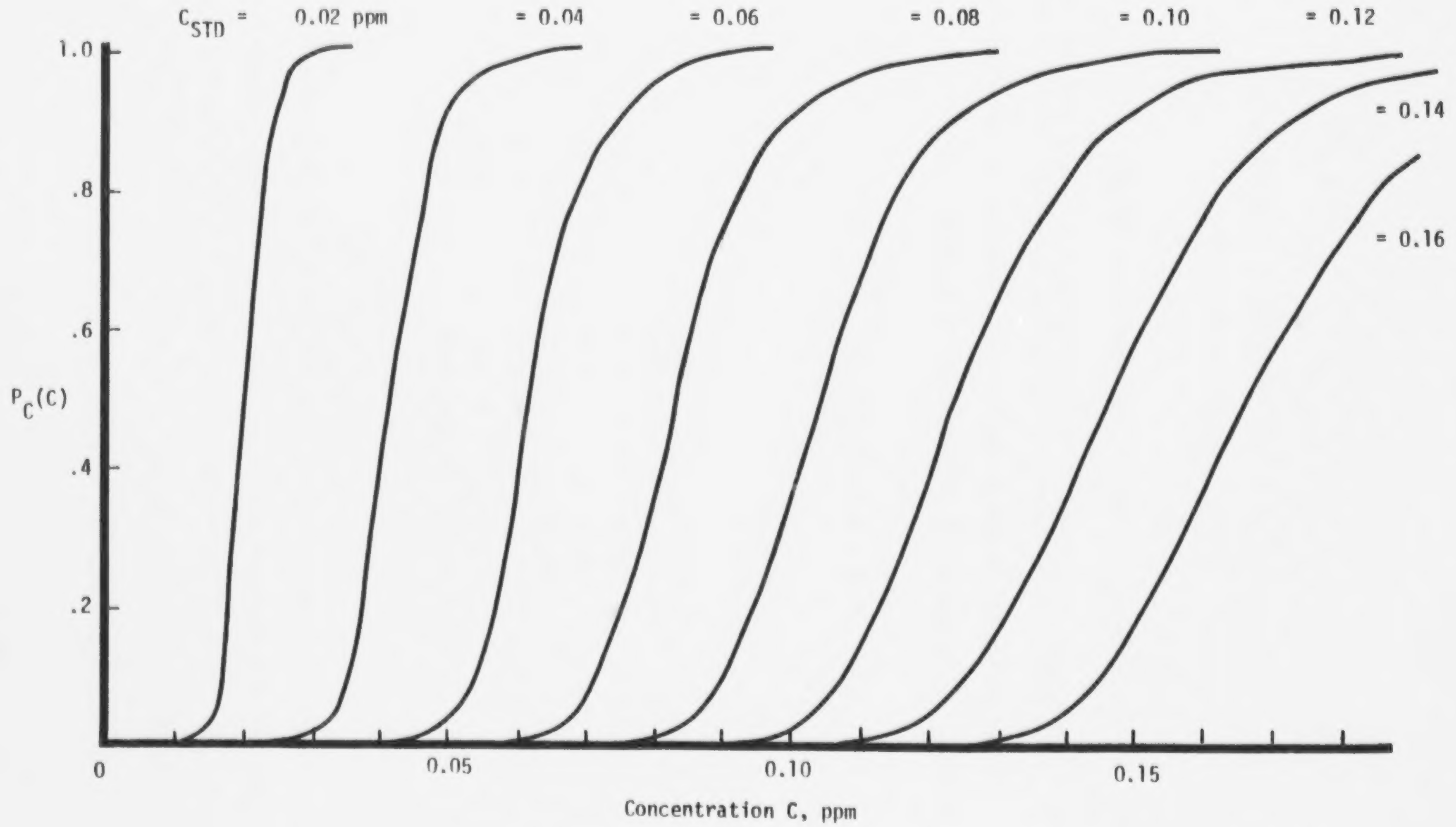
$$P_C(C) = \left[1 - \exp\left(\frac{-\ln 8760}{C_{STD}} C\right)\right]^{8760} \qquad \text{(D-1)}$$

Thus:

C_{STD}	$\ln 8760/C_{STD}$	C_{STD}	$\ln 8760/C_{STD}$
0.06	151.30	0.13	69.83
0.07	129.69	0.14	64.84
0.08	113.47	0.15	60.52
0.09	100.87	0.16	56.74
0.10	90.78	0.17	53.40
0.11	82.53	0.18	50.433
0.12	75.65		

Substituting a constant from this table into Eq. (D-1) above gives the $P_C(C)$ function corresponding to each level of the standard. Figure D-1 shows plots of $P_C(C)$ for various levels of the standard generated from the constants in the table.

Figure D-1 Change in Function $P_C(C)$ with Change in C_{STD} for Exponential Distribution



D.2 Form the Composite Distribution Function, $p^\circ_T(C)$, For the Three Thresholds

Equation (4-5), Section 4.2 is used to calculate the composite distribution function. In the example, theree health effects are considered. Expanding Eq. (4-5) for three thresholds yields:

$$p^\circ_T(C) = e_1Q_2Q_3p^\circ_{T_1} + e_2Q_1Q_3p^\circ_{T_2} + e_3Q_1Q_2p^\circ_{T_3} \qquad \text{(D-2)}$$

Note from Eq. (4-6) that:

$$Q_i = 1 - e_iQ^\circ_i \qquad \text{(D-3)}$$

where:

$$Q^\circ_i = \int_0^C p^\circ_i \, dC \qquad \text{(D-4)}$$

Thus, in addition to requiring tabulations of the threshold probability density functions p°_i versus concentration, tabulations of the cumulative distribution is also needed. In the example all three p°_i are normal distributions.

From the constants provided in the small table on the first page of Section 4.5 one can write:

$$p^\circ_1 = \frac{1}{2\text{x}0.05\ \sqrt{2\pi}} \exp\left(-(C-0.15)^2 / 2 \times (0.05)^2\right)$$

$$p^\circ_i = 3.989 \exp\left(-200\ (C-0.15)^2\right) \qquad \text{(D-5)}$$

The functions for p°_2 and p°_3 can be developed in the same manner.

TABLE D-1

INDIVIDUAL AND COMPOSITE HEALTH EFFECTS THRESHOLD PROBABILITY DENSITY FUNCTIONS VS. CONCENTRATION

STANDARD (ppm)	$p^{\circ}_{T_1}$	$p^{\circ}_{T_2}$	$p^{\circ}_{T_3}$	Q°_1	Q°_2	Q°_3	p°_{MT}	$p^{\circ}_{1,3}$
0.00	0.09	0.00	0.00	0.001	0.000	0.000	0.01	0.01
0.01	0.16	0.00	0.01	0.003	0.000	0.000	0.03	0.03
0.02	0.27	0.00	0.01	0.005	0.000	0.000	0.05	0.05
0.03	0.45	0.00	0.02	0.008	0.000	0.000	0.09	0.09
0.04	0.71	0.00	0.03	0.014	0.000	0.001	0.14	0.14
0.05	1.08	0.01	0.04	0.023	0.000	0.001	0.21	0.20
0.06	1.58	0.02	0.06	0.036	0.000	0.002	0.31	0.30
0.07	2.22	0.05	0.08	0.055	0.000	0.002	0.45	0.41
0.08	2.99	0.11	0.11	0.081	0.001	0.003	0.64	0.56
0.09	3.88	0.23	0.16	0.115	0.003	0.004	0.92	0.74
0.10	4.84	0.44	0.22	0.159	0.006	0.006	1.27	0.94
0.11	5.79	0.79	0.30	0.212	0.012	0.009	1.77	1.15
0.12	6.66	1.35	0.40	0.274	0.023	0.012	2.37	1.37
0.13	7.37	2.16	0.52	0.345	0.040	0.017	3.14	1.58
0.14	7.82	3.24	0.67	0.421	0.067	0.023	4.05	1.77
0.15	7.98	4.57	0.86	0.500	0.106	0.030	5.07	1.96
0.16	7.82	6.05	1.08	0.579	0.159	0.040	6.09	2.11
0.17	7.37	7.53	1.33	0.655	0.227	0.052	6.99	2.25
0.18	6.66	8.80	1.62	0.726	0.308	0.067	7.64	2.38
0.19	5.79	9.67	1.94	0.788	0.401	0.100	7.83	2.49
0.20	4.84	9.97	2.28	0.841	0.500	0.106	7.82	2.64
0.21	3.88	9.67	2.65	0.885	0.599	0.129	7.30	2.81
0.22	2.99	8.80	3.02	0.919	0.692	0.159	6.44	2.98
0.23	2.22	7.53	3.40	0.945	0.773	0.189	5.41	3.19
0.24	1.58	6.05	3.76	0.964	0.841	0.227	4.31	3.40
0.25	1.08	4.57	4.10	0.977	0.894	0.264	3.33	3.62
0.26	0.71	3.24	4.40	0.986	0.933	0.309	2.50	3.82

Table D-1 (CONTINUED)

STANDARD (ppm)	$P^\circ_{T_1}$	$P^\circ_{T_2}$	$P^\circ_{T_3}$	Q°_1	Q°_2	Q°_3	P°_{MT}	$P^\circ_{1,3}$
0.27	0.45	2.16	4.65	0.992	0.956	0.352	1.89	4.00
0.28	0.27	1.35	4.83	0.995	0.977	0.401	1.45	4.13
0.29	0.16	0.79	4.95	0.997	0.988	0.448	1.18	4.22
0.30	0.09	0.44	4.99	0.999	0.994	0.500	1.02	4.25
0.31	0.05	0.23	4.95	1.000	0.997	0.552	0.92	4.21
0.32	0.02	0.11	4.83		0.999	0.599	0.85	4.11
0.33	0.01	0.05	4.65		1.000	0.648	0.80	3.95
0.34	0.01	0.02	4.40			0.691	0.75	3.74
0.35	0.00	0.01	4.10			0.736	0.70	3.49
0.36	0.00	0.00	3.76			0.773	0.64	3.20
0.37	0.00	0.00	3.40			0.811	0.58	2.89
0.38	0.00	0.00	3.02			0.841	0.51	2.57
0.39	0.00	0.00	2.65			0.871	0.45	2.25
0.40	0.00	0.00	2.28				0.39	1.94
0.41			1.94			0.900	0.33	1.65
0.42			1.62				0.28	1.38
0.43			1.33			0.948	0.23	1.13
0.44			1.08				0.18	0.92
0.45			0.86			0.970	0.15	0.73
0.46			0.67				0.11	0.57
0.47			0.52			0.983	0.09	0.44
0.48			0.40				0.07	0.34
0.49			0.30			0.991	0.05	0.26
0.50			0.22				0.04	0.19
0.51			0.16			0.996	0.03	0.14
0.52			0.11				0.02	0.09
0.53			0.09			0.998	0.01	0.07
0.54			0.06				0.01	0.05
0.55			0.04			0.999	0.01	0.03
0.56			0.03				0.01	0.03
0.57			0.02			1.000	0.00	0.02
0.58			0.01					0.01
0.59			0.01					0.01
0.60			0.00					0.00

Since the distributions are normal, their integrals as a function of $(C-\bar{C}/\sigma)$ can be obtained from tables generally available in statistical textbooks. The p_i functions and integrals, Q_i°, are tabulated for various values of C in Table D-1 of this appendix. The composite distribution p_T° and the semicomposite distribution $p_{1,3}$ are also given in the table. The function p_T° is calculated from the individual $p_{T_i}^\circ$ as follows: Take as an example C = 0.20 ppm. By substitution of the value at this concentration calcualte the Q_i (0.20) from Eq. (D-3).

$$Q_1\ (0.20) = 1 - 0.15 \times 0.841 = 0.874$$

$$Q_2\ (0.20) = 1 - 0.80 \times 0.500 = 0.600$$

$$Q_3\ (0.20) = 1 - 1.0 \times 0.106 = 0.894$$

Note the e_1 are from the table on the first page of Section 4.5. Substitutin these values and the p_i° (0.20) from the table into Eq. (D-2) gives:

$$p_{MT}^\circ\ (0.20) = 0.15 \times 0.6 \times 0.894 \times 4.84 + 0.8 \times 0.874\ 0.894 \times 9.97 + 1.0 \times 0.874 \times 0.600 \times 2.28 = 7.82$$

The function $p_{1,3}$ is calculated in a similar manner.

D.3 Determine the Risk, R, That no True Threshold Will be Exceeded

The risk is calculated from Eq. (4-2). Note that the product of probabilities terms in Eq. (4-4) is zero because $e_3 = 1$. For this case, therefore, one can write:

$$R = 1 \int_0^\infty P_C\ (C, C_{STD})\ p_T^\circ\ (C) dC. \qquad \text{(D-5)}$$

Given any potential value of the standard, the appropriate constant can be calculated for P_C as was done in the Table in Step #1. Substituting this constant into Eq. (C-1) a value of P_C to be calculated for every value of p°_T in Table D-1 of this appendix. The corresponding product is then calculated and numerically integrated over the range of p°_T.

For example, for a level of the standard C_{STD} = 0.14 ppm have:

$$P_C(C) \;=\; [1 - \exp(-64.84C)]^{8760}$$

Evaluating this expression at C = 0.20 ppm gives:

$$P_C(0.2) = [1 - \exp(-64.84 \times 0.2)]^{8760} = 0.980$$

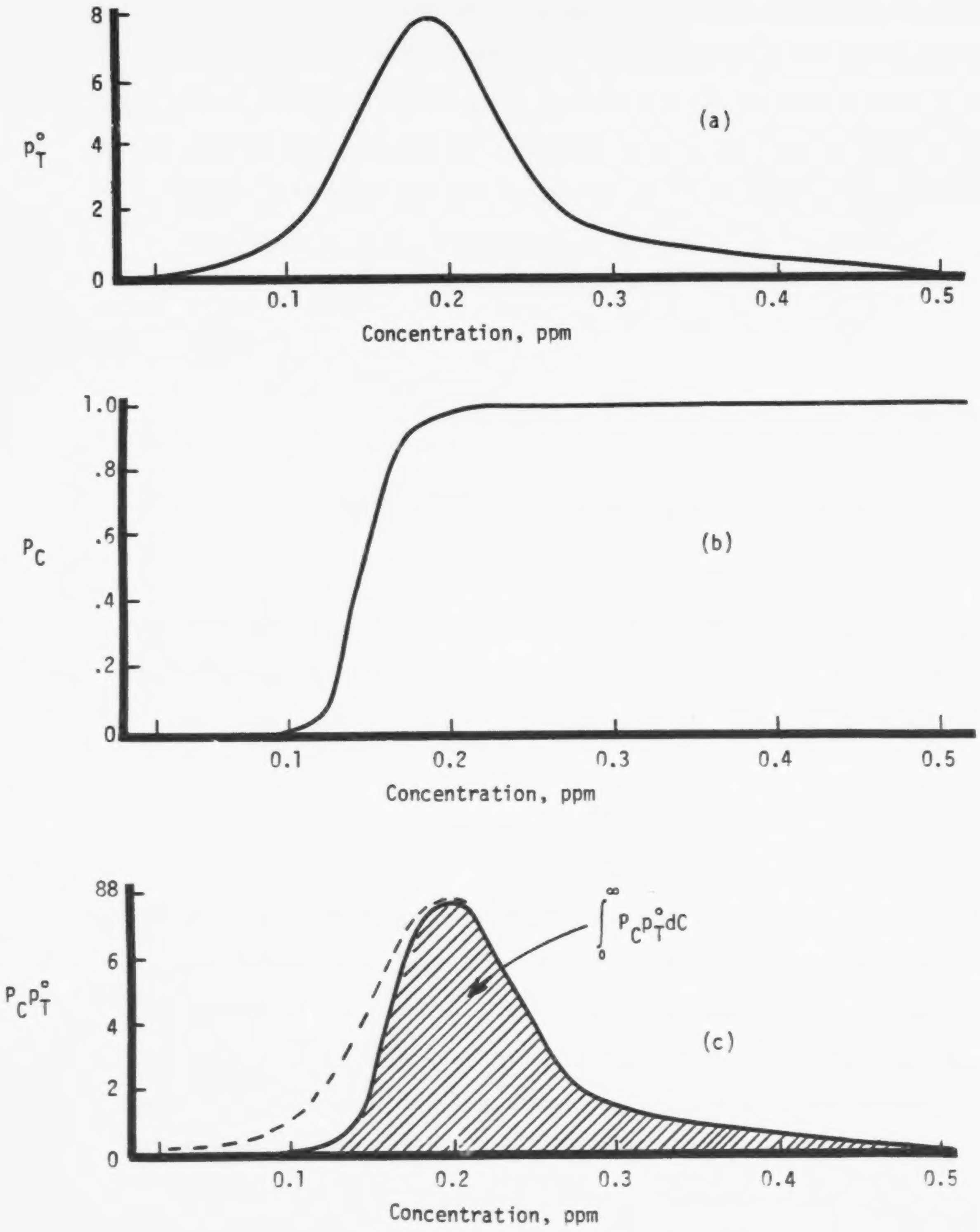
That is, there is a 0.980 chance that the hourly average 0.2 ppm will not be exceeded in a given calendar year. The corresponding p°_T was calculated in Step #3 to be 7.82. The product is:

$$P_C\,(0.2)\;p^\circ_T\,(0.2) = 7.66$$

This product is tabulated as concentration over the range of p°_T. The product is then numerically integrated over the range of p°_T using methods such as Simpsons' rule or trapezoidal integration. The procedure is repeated for each level of the standard shown in Table 4-4, Section 4.5.

Figure D-2 shows the procedure graphically for a standard at 0.14 ppm. Figure D-2(a) shows the P_C function. Figure D-2(b) shows the composite probability density function p°_T, and Figure D-2(c) shows their product.

Figure D-2 Graphical Illustration of p_T Calculation



The dashed line in Figure D-2(c) sketches in the original composite density function. The area under the solid curve is p_T, the probability of not exceeding the threshold. The area between the dashed line and the full line is the probability of exceeding the true threshold. The sum of the areas, is of course, the total area under the composite density function alone.

D.4 Determine R_i

The risk, R_i, of exceeding any individual threshold is determined proceeding in the same manner as described in Step #3.

D.5 Determine any $R_{i,j}$

The risk of exceeding the lower of two thresholds, $R_{i,j}$ can be calculated by forming the composite in a manner similar to Step #2 and then proceeding to Step #3.